TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Petri Helin

Wideband Direction of Arrival estimation and sparse modeling for underwater surveillance

Master's thesis

Examiners and topic approved by the
Faculty Council on 8 May 2013.
Examiners: Professor Ioan Tabus
Professor Jaakko Astola

Abstract

Tampere University of Technology
Master's Degree Programme in Information Technology
Helin, Petri: Wideband Direction of Arrival estimation and sparse modeling for underwater surveillance
Master's thesis: 97 pages
August 2013
Major subject: Signal Processing
Examiners: Professor Ioan Tabus and Professor Jaakko Astola
Keywords: Underwater Surveillance, Passive Sonar, Wideband DOA estimation, Sparse Modeling, MVDR-LBC

In underwater surveillance sources, such as ships or submarines, are localized using the acoustic noise emitted by the source engines, propellers and other machinery. The acoustic signals propagate in the sea and are recorded with an array of acoustic sensors. Processing the recorded signals to obtain the locations of the sources is known as Direction of Arrival (DOA) estimation in the field of signal processing.

A simple mathematical model relating the sensor array geometry to the DOA of the source exists when the frequency of the source signal is known. The model is directly applicable to a narrowband DOA estimation problem where the energy of the source signals is concentrated around a single carrier frequency. For underwater surveillance, however, the source signals are wideband which complicates the problem.

This thesis reviews existing methods for wideband DOA estimation: Simple extensions of well known narrowband methods MVDR and MUSIC, the so called coherent methods and the most recent methods belonging into the sparse framework. An original idea for extending MVDR using a likelihood based combining of subbands, MVDR-LBC is developed.

The thesis models the sensor signals as a sparse autoregressive process by linear prediction and the original algorithm GRLS. The sparse model is shown to be effective compared to the conventional non-sparse one. The model can be used to compress the data recorded in underwater surveillance.

The wideband DOA estimation methods are tested with a number of simulations and with real data recorded in the sea. MVDR is shown to be robust and effective, the accuracy and resolution of which can be improved using MVDR-LBC. MUSIC provides good resolution, is computationally efficient and can be implemented quite simply. The coherent methods are the most complicated and need good preestimations for the source directions but can resolve close sources best.

Preface

This is a Master of Science thesis written for the Department of Signal Processing in Tampere University of Technology. The research work, that is covered in the thesis, revolved around the topic of *underwater surveillance*. During the last decade, the mathematical tools relating to the so-called *sparsity* have attracted a lot of interest in the field of signal processing and they were also an inspiration to this work. They proved out useful when modeling the signals encountered in underwater surveillance with an original sparse modeling method using the algorithm GRLS. An original algorithm called MVDR-LBC that improves a conventional, robust DOA estimation method was also invented during the research work.

The work was supervised by Professor Ioan Tabus who shared his expertise in signal processing and mathematics by inspiring, sharing ideas and giving technical and practical advice. He helped in developing MVDR-LBC and the method for sparse modeling and helped me to finalize the thesis. All the same can also be said about D.Sc. Bogdan Dumitrescu who is also the main inventor of GRLS. I also want to thank Professor Jaakko Astola who shared his helpful and inspirational views on DOA estimation and M.Sc. Alexandru Onose who is one of the inventors of GRLS and also gave some DOA estimation ideas in our meetings.

August 12, 2013

Petri Helin
petri.helin@gmail.com

Contents

Glossaries

Glossary of terms and abbreviations.

Term	Explanation
AIC	Akaike Information Criterion. An information theoretic criterion.
AR	Autoregressive.
CSSM	Coherent Signal Subspace Method. A wideband DOA estimation method.
DFT	Discrete Fourier Transform.
DOA	Direction of Arrival.
DOI	Direction of Interest, the look out angle.
EVD	Eigenvalue decomposition.
FFT	Fast Fourier Transform.
FIR	Finite Impulse Response filter.
GRLS	Greedy Sparse RLS. A sparse adaptive filtering algorithm.
IIR	Infinite Impulse Response filter.
MDL	Minimum Description Length. An information theoretic criterion.
RLS	Recursive Least Squares. An adaptive filtering algorithm.
SNR	Signal to noise ratio.
Source	A noise emitting vessel such as a submarine or a ship.
SVD	Singular value decomposition.
Subband	A narrowband component of the decomposed wideband signal.
MUSIC	Multiple Signal Classification. A DOA estimation method.
MUSIC-S	A wideband DOA estimation method based on MUSIC and summation.
MVDR	Minimum Variance Distortioneless Response. A DOA estimation method.
MVDR-LBC	A wideband DOA estimation method based on MVDR and the likelihood function.
MVDR-S	A wideband DOA estimation method based on MVDR and summation.
OMP	Orthogonal Matching Pursuit. A sparse recovery algorithm.
OOMP	Optimized Orthogonal Matching Pursuit. A sparse recovery algorithm.
TOPS	Test of Orthogonality of Projected Subspaces. A wideband DOA estimation method.
WAVES	Weighted Average of Signal Subspaces. A wideband DOA estimation method.

Glossary of mathematical notation.

Operator	Explanation
$\boldsymbol{x}$	vectors and matrices are denoted by a bold face letter
^{T}	transpose
^{H}	hermitian transpose
*	the complex conjugate
$\mathrm{Re}\left[x\right]$	the real part of x
$\mathrm{E}\left[x\right]$	the expectation of the random variable x
$\hat{x}$	an estimate of x
$\dot{x}$	the derivative of x
$\ddot{x}$	the second order derivative of x
$\left\|\cdot\right\|_2$	the euclidean norm
$\left\|\cdot\right\|_{\ell_p}$	the ℓ_p norm
$\left\|\cdot\right\|_F$	the Frobenius norm

1 Introduction

In the time of World War I, under the increasing threat of enemy submarines, the first devices for locating the enemy vessels by the noise caused by their engines and propellers were developed. These early passive sonar devices consisted of sound capturing tubes attached to stethoscope earpieces. By rotating the tubes, an operator could listen when the sounds of the earpieces were equal. This rotation corresponded to the angle in which the vessel – the source – was located. A modern day equivalent of this kind of underwater surveillance system consists of an unmanned underwater vehicle, UUV, submerged in the sea and an array of sensors towed by the UUV. The towing is done so that the own noise of the UUV would not mask the incoming signals from other vessels. As the name suggests, a sensor array consists of a number of acoustic sensors i.e. hydrophones.

A major part of this thesis is the automatic process of using signal processing to locate the signal emitting sources which is referred to as Direction of Arrival (DOA) estimation. Saving and transmitting the recorded signals in the UUV is another practical problem encountered in underwater surveillance. Another part of the thesis is to model the received acoustic signals using a sparse autoregressive process which can be utilized to compress the signals.

The field of wideband DOA estimation is of interest to us because in underwater surveillance the source signals have a wide frequency spectrum. The machinery causing the noise emitted by them can range from a few dozens of hertz to several kilohertz. This is different to other fields of DOA estimation such as radar, active sonar and telecommunications where the source signal can be assumed to have a narrowband frequency spectrum which means that the frequency range of the signal is small compared to the carrier frequency.

In order to perform the DOA estimation, a model relating the locations of the sources to the measurements of the sensor array is needed. Basically the idea is that all the sensors receive the same signal at slightly different time instants. The delay in time relates to the direction of the source. The very simplest method for wideband DOA estimation called Delay-and-Sum is based on this simple idea: it uses a hypothesized angle to delay the signals correspondingly and then sums up the signals. This is applicable to both, narrow and wideband signals and is in a way a counterpart of the analog equipment used in the World War I era.

We will see, that when it comes to narrowband signals the DOA signal model can be greatly simplified. Since the phase depends only on the DOA, a simple multiplication with an exponential is able to compensate for the delay. This model is analogous to the sinusoidal signal model in the field of frequency spectrum estimation. The simple sinusoidal model does, however, not directly apply to wideband signals since

the frequency also affects the phase. Yet, most wideband DOA estimation methods still rely on the narrowband model on one level or another. They operate by decomposing the wideband signal into many narrowband signals using a filter bank. The *incoherent* methods process the narrowband problems individually and then combine the results. The *coherent* methods, meanwhile, use the narrowband signals to compute estimates for e.g. covariance matrices but eventually perform the estimation on a single model describing the whole wideband signal.

The thesis describes three classical narrowband DOA methods. The Beamformer is the spatial equivalent of the periodogram in frequency estimation and as such acts as a reference method. Superresolution methods that can outperform the Beamformer are Capon's method, which is also known as MVDR, and MUltiple SIgnal Classification, MUSIC. MVDR became known as early as 1969 being introduced by Capon in [1] and MUSIC followed ten years later introduced by Schmidt in [2]. They are conventionally used for the wideband problem via a simple additive combining over the frequency bands. We, however, devoloped a novel idea, that is published in [3], about using MVDR for the wideband problem and name it MVDR-LBC.

The mentioned coherent wideband DOA methods originate from a method called CSSM that was originally presented as early as 1988 by Hung and Kaveh in [4]. The coherent methods have seen some development in the last decade with methods such as WAVES [5] and TOPS [6]. All these methods are presented in the thesis.

This work was originally motivated by the most recent DOA estimation methods which are based on the recent hot topic in signal processing – sparsity. The first sparse method for DOA estimation based on ℓ_1-minimization, namely l1-SVD, was introduced by Malioutov in [7]. In general, sparse problems are often solved with more tractable greedy algorithms which have been used for wideband DOA estimation in [8]. Also, other articles such as [9] and [10] concerning greedy algorithms have been published. The applicability of the sparse methods to the wideband DOA problem is covered in the thesis but we also cover another aspect of sparsity and use an original idea based on the publication [11] to model the underwater signals.

The DOA estimation methods are required to satisfy certain requirements such as the ability to resolve close sources, to be robust to unidealities and to provide accurate estimates. The performance of the methods is analyzed through experiments using simulated and real data.

The structure of the thesis is the following. First, the signal model for DOA estimation is developed in Section 2. Using the model, it is shown how the narrowband DOA estimation methods found in the literature work in Section 3. Section 4 covers the main part of the thesis. It describes existing wideband DOA methods which typically are extensions of the narrowband methods. In Section 5, the original method

for wideband DOA estimation is introduced and described. Sparse modeling for the underwater signals is covered in Section 6. In Section 7, the measurement system and the signals encountered in underwater surveillance are discussed. Experimental results for wideband DOA estimation and sparse modeling are given in Section 8 and finally the thesis is concluded in Section 9.

2 Signal model for DOA estimation

The basic device needed for DOA estimation is called a *sensor array*. In an underwater environment it is an array of hydrophones. In order to estimate the DOA of a single source, the sensors should receive the same signal but at slighlty different time instants. This is accomplished by placing them into different positions in space. Figure 1 shows a very typical configuration of a sensor array – a uniform linear array, *ULA* – where the sensors are positioned along the same line at equal distances from each other. The angle θ is the *direction of arrival*. This thesis is all about estimating θ.

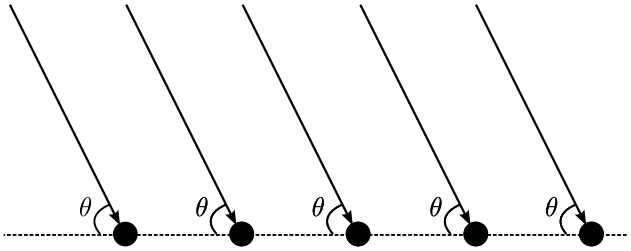


Figure 1: A uniform linear array. The black dots indicate hydrophones and the arrows the signal impinging on them.

The signal arriving at the sensors in Figure 1 comes from a single source. This signal measured at a reference point is denoted by $s(\tau)$[1]. The distance from the source to the hydrophones is much larger than the spacing between the hydrophones which is why the rays can be considered parallel and the angle θ same for all sensors. This is generally known as the far-field assumption.

Let $\Delta\tau_n$ denote the time that the signal takes to propagate from the reference point to the n-th sensor. The signal received at the n-th sensor can be written

$$u_n(\tau) = s(\tau - \Delta\tau_n) + e_n(\tau) \tag{1}$$

where $e_n(\tau)$ is the additive noise term.

Figure 2 shows a closer view of the first three hydrophones in a linear array. They are set into the space along the same line so that the distance from the first sensor to the second is x_2. The same is true for other hydrophones, the total number of which is N. From Figure 2 it can be said that the extra distance the signal has to travel from the first hydrophone to the second is

$$d_2 = x_2 \cos\theta. \tag{2}$$

The delay in time is the distance divided by the speed of propagation (in our case,

[1]The continuous time is denoted by τ while the discrete time instant (sample) is denoted by t.

the speed of sound in water) so the propagation time to the second sensor obeys the expression

$$\Delta\tau_2 = \Delta\tau_1 + \frac{d_2}{c} = \frac{x_2 \cos\theta}{c}. \tag{3}$$

In general we can say similarly that

$$\Delta\tau_n = \Delta\tau_1 + \frac{x_n \cos\theta}{c}$$

for $n = 1, \ldots, N$. Since $\Delta\tau_1$ is common for all $\Delta\tau_n$ we can treat the first sensor as the reference point which implies $\Delta\tau_1 = 0$ and in general:

$$\Delta\tau_n = \frac{x_n \cos\theta}{c} \tag{4}$$

for $n = 1, 2, \ldots, N$.

Combining (1) and (4) gives

$$u_n(\tau) = s\left(\tau - \frac{x_n \cos\theta}{c}\right) + e_n(\tau) \tag{5}$$

The estimation of θ from a single source essentially boils down to (5). It is true for both wideband and narrowband signals. A model for wideband sources can not be simplified as such but for narrowband signals it is easy to delay the signal with a simple exponential multiplication. This narrowband model can also be utilized when processing wideband signals and is presented next.

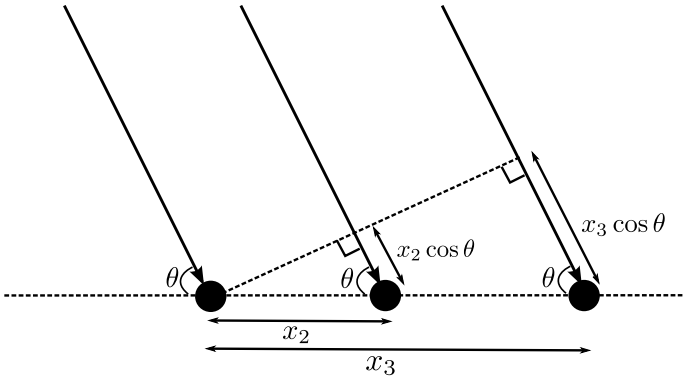


Figure 2: The first three hydrophones in the sensor array. x_n is the distance from the first to n'th sensor.

2.1 Single narrowband source

A narrowband signal has its energy concentrated around a central frequency ω_c. Such signals are encountered as such in telecommunications and radar systems for example. To us they are interesting because the wideband signal is typically decom-

posed into narrowband signals. For mathematical simplicity, the signals are treated as complex valued. Finding the complex presentation for the received signal with a hardware demodulator is depicted in [12, p. 270]. This is, however, not important for us since the wideband signals will obtain their complex narrowband form naturally through the use of FFT.

A truly narrowband signal consists of the information containing signal which is modulated around a high carrier frequency ω_c. The information signal has the expression $\alpha(t)e^{j\phi(t)}$ and it should have a narrow baseband i.e. it only contains power over low frequencies. α is the amplitude and ϕ the phase of the signal.

The modulated signal

$$s(\tau) = \alpha(\tau)e^{j\omega_c\tau}e^{j\phi(\tau)} \tag{6}$$

is received at the reference point. A hydrophone receives the signal delayed by $\Delta\tau$:

$$s(\tau - \Delta\tau) = \alpha(\tau - \Delta\tau)e^{j\omega_c(\tau-\Delta\tau)}e^{j\phi(\tau-\Delta\tau)} = \alpha(\tau - \Delta\tau)e^{j\omega_c\tau}e^{j\phi(\tau-\Delta\tau)}e^{-j\omega_c\Delta\tau}. \tag{7}$$

The full derivation for the following model equation through the frequency domain is given in [12, pp. 268–271]. We can, however, give some intuitive reasoning to it by considering that if $\Delta\tau$ is small the information containing signal stays almost constant since it was only containing low frequency components and so $\alpha(\tau-\Delta\tau) \approx \alpha(\tau)$ and $\phi(\tau - \Delta\tau) \approx \phi(\tau)$ which simplifies (7) to

$$s(\tau - \Delta\tau) = s(\tau)e^{-j\omega_c\Delta\tau}. \tag{8}$$

By giving $\Delta\tau$ the delay corresponding to a hydrophone, (5) simplifies to

$$u_n(\tau) = s(\tau)e^{-j\omega_c\Delta\tau_n} + e_n(\tau). \tag{9}$$

By combining all the sensors into a vector we have

$$\begin{bmatrix} u_1(\tau) \\ u_2(\tau) \\ \vdots \\ u_N(\tau) \end{bmatrix} = \begin{bmatrix} e^{-j\omega_c\Delta\tau_1} \\ e^{-j\omega_c\Delta\tau_2} \\ \vdots \\ e^{-j\omega_c\Delta\tau_N} \end{bmatrix} s(\tau) + \begin{bmatrix} e_1(\tau) \\ e_2(\tau) \\ \vdots \\ e_N(\tau) \end{bmatrix} \tag{10}$$

where the vector

$$\boldsymbol{a}(\theta, \omega_c) = \begin{bmatrix} e^{-j\omega_c\Delta\tau_1} & e^{-j\omega_c\Delta\tau_2} & \cdots & e^{-j\omega_c\Delta\tau_N} \end{bmatrix}^T \tag{11}$$

is called the *steering vector* and $\Delta\tau_n$ is given by (4).

The equation (10) is the basic model for DOA estimation when the source signal is narrowband.

2.2 Decomposing the wideband signal

In the case of underwater surveillance or *passive sonar*, as the field is sometimes referred to in the literature, the signals are wideband. For wideband signals it is not possible to build a direct model as simple as (10) due to the steering vector of (11) being dependent on the frequency ω_c. For wideband signals no such unique frequency exists! The basic approach for dealing with wideband sources is to just decompose the wideband signals into several narrowband signals by means of a filter bank.

Before using the filter bank, the analogue signals $u_n(\tau)$ are sampled at the sampling rate f_s. From now on, we use the same symbol u_n to denote the discretized signals. The sample index is t. The discretized signals are passed into the filter bank. The most common way of constructing a filter bank in DOA estimation is the discrete Fourier transform (DFT) which is simple and fast to implement using the fast Fourier transform (FFT). The filter bank interpretation of DFT is depicted in Figure 3. Each branch of it consists of a complex modulated prototype filter $H(z)$ and a decimation block producing the narrowband output signals y_n. The amplitude responses of the filters in the filter bank are illustrated in Figure 4 with 32 frequency bands. Although the responses are not very ideal since the sidelobes are quite strong, the simple DFT filter bank is considered sufficient since the DOA estimation is anyway done in a very high noise environment and the filtering part should not play a very critical role.

N_{fft} samples are taken from the time domain signal $u_n(t)$ to compute the Fourier transform. In this way the frequency space from 0 to f_s is covered such that the k-th frequency is

$$f_k = \frac{k-1}{N_{\text{fft}}} f_s \qquad k = 1, \ldots, K \tag{12}$$

where $K = N_{\text{fft}}/2$ since the other half of the frequency space is only the complex conjugate of the first half and essentially carries the same information. In order to have several time samples (snapshots) for each narrowband signal we take P consecutive frames of length N_{fft} and proceed by computing the DFT. Thus we have the transformed signals

$$y_n(p, \omega_k) = \frac{1}{\sqrt{N_{\text{fft}}}} \sum_{t=0}^{N_{\text{fft}}-1} h(t) u_n\left((p-1)N_{\text{fft}} + t\right) e^{-j\omega_k t} \qquad p = 1, \ldots, P \tag{13}$$

where $h(t)$ is the window function, p is the snapshot index and $\omega_k = 2\pi f_k / f_s$. The Fourier transform is applied to the product of the windowing function $h(t)$ and the input signal $u_n(t)$. As can be seen from the scaling coefficient, the Fourier transform is chosen to be symmetric (to the inverse transform) in order to conserve the signal

power i.e.

$$\sum_{t=0}^{N_{\text{fft}}-1} |h(t)u_n((p-1)N_{\text{fft}}+t)|^2 = \sum_{k=1}^{N_{\text{fft}}} |y_n(p,\omega_k)|^2 \tag{14}$$

The filter $H(z)$ in the filter bank (figure 3) corresponds to the window function $h(t)$, which is chosen to be the rectangular window.

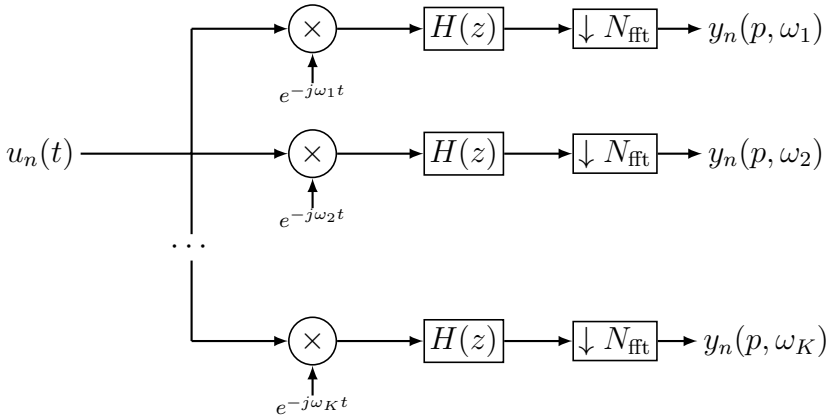


Figure 3: DFT filter bank. The exponential multiplication acts as a demodulation i.e. it moves the frequency at ω_k to the zero frequency. $H(z)$ is a lowpass filter.

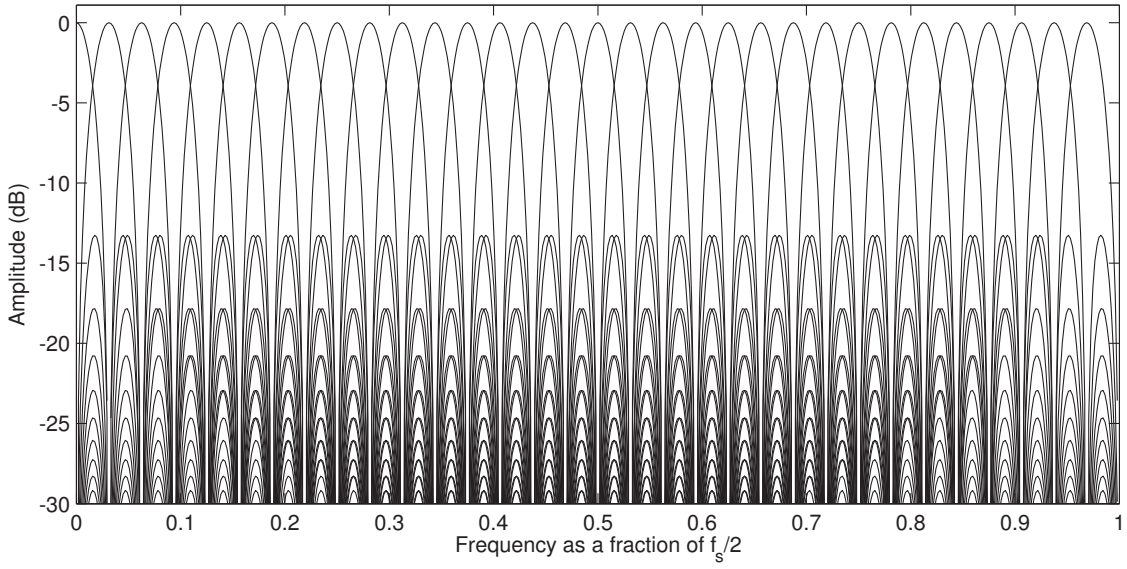


Figure 4: The filter bank for $K = 32$.

2.3 Multiple wideband sources

The final model for wideband DOA estimation is built using the results in the previous sections. The fact that signals from multiple sources can impinge on the sensor array simultaneously can be taken into account by a simple superposition principle.

The filter bank outputs narrowband signals of similar form as in Section 2.1. The model described by equation (10) now exists for all frequency bands. Now we also consider that there are Q sources instead of one. The received signal is a superposition of the source signals. Let Q denote the number of sources and $s_q(p, \omega_k)$ the narrowband component corresponding to the center frequency ω_k of the q-th source at the snapshot index p. The model equation (10) can then be written as

$$y_n(p, \omega_k) = a_n(\theta_1, \omega_k)s_1(p, \omega_k) + a_n(\theta_2, \omega_k)s_2(p, \omega_k) + \cdots + a_n(\theta_Q, \omega_k)s_Q(p, \omega_k) + e_n(p, \omega_k) \quad (15)$$

and gathering all the sensors into a vector

$$\begin{bmatrix} y_1(p, \omega_k) \\ y_2(p, \omega_k) \\ \vdots \\ y_N(p, \omega_k) \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}(\theta_1, \omega_k) & \boldsymbol{a}(\theta_2, \omega_k) & \cdots & \boldsymbol{a}(\theta_Q, \omega_k) \end{bmatrix} \begin{bmatrix} s_1(p, \omega_k) \\ s_2(p, \omega_k) \\ \vdots \\ s_Q(p, \omega_k) \end{bmatrix} + \begin{bmatrix} e_1(p, \omega_k) \\ e_2(p, \omega_k) \\ \vdots \\ e_N(p, \omega_k) \end{bmatrix}. \quad (16)$$

The same system of equations which is very essential and frequently used in this thesis can be expressed in short as:

$$\boldsymbol{y}(p, \omega_k) = \boldsymbol{A}(\omega_k)\boldsymbol{x}(p, \omega_k) + \boldsymbol{e}(p, \omega_k) \qquad p = 1, \ldots, P \quad k = 1, \ldots, K. \quad (17)$$

The steering vectors don't depend on p so the matrix containing steering vectors as columns is the same for all snapshots. This is why the snapshots of $\boldsymbol{y}$, $\boldsymbol{s}$ and $\boldsymbol{e}$ can be arranged as columns to arrive at a single matrix equation for all snapshots:

$$\boldsymbol{Y}(\omega_k) = \boldsymbol{A}(\omega_k)\boldsymbol{X}(\omega_k) + \boldsymbol{E}(\omega_k) \qquad k = 1, \ldots, K \quad (18)$$

where $\boldsymbol{Y}$ is $N \times P$, $\boldsymbol{A}$ is $N \times Q$, $\boldsymbol{X}$ is $Q \times P$ and $\boldsymbol{E}$ is $N \times P$. Typical exemplar values are $N = 16$, $P = 64$, $K = 64$ and $Q < N$.

3 Narrowband DOA methods

Many signals such as those encountered in radar or sonar applications or in telecommunications have narrow frequency spectra. Thus, estimating the direction of arrival in one narrowband is sufficient. Narrowband estimation methods are, however, also a fundamental part in many wideband DOA estimation methods which is why the most common are also described in this section.

We begin by describing the most basic narrowband DOA estimation method, namely the *Beamformer*. According to [12] it can be viewed as a counterpart of the temporal filter bank method which is also known as the periodogram. The idea of spatial filtering also applies to Capon's method, MVDR.

The parametric temporal methods such as MUSIC and Esprit can also be used in the spatial domain. We describe MUSIC in detail.

3.1 Spatial Filtering

Classical methods aim to design a filter that receives a signal undistortioned in one direction (angle θ) while attenuating all the other directions. Then the power of the received signal in that direction is calculated. The directions where the power estimate peaks define the DOA estimates. Spatial filtering is illustrated in Figure 5.

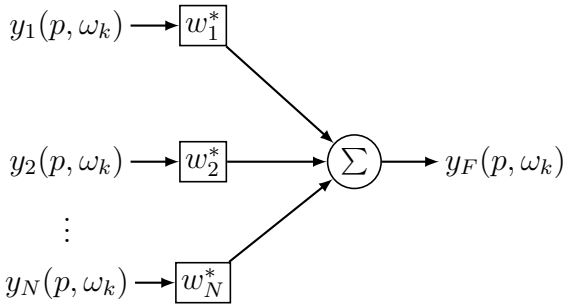


Figure 5: Spatial filtering. The signal received at the sensors is multiplied by the filter coefficients. This is analogous to convolution in time domain.

The vector of signals received at different hydrophones $[y_1, y_2, \ldots, y_N]^T$ is filtered with $\boldsymbol{w}$ to obtain y_F. As an equation this means:

$$y_F(p,\omega_k) = \boldsymbol{w}^H \boldsymbol{y}(p,\omega_k). \tag{19}$$

Combining equation (19) with the signal model (17) (we drop the snapshot and

frequency indices for simplicity):

$$\begin{aligned} y_F &= \boldsymbol{w}^H(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}) \\ &= \boldsymbol{w}^H\boldsymbol{a}(\theta_1)s_1 + \boldsymbol{w}^H\boldsymbol{a}(\theta_2)s_2 + \cdots + \boldsymbol{w}^H\boldsymbol{a}(\theta_Q)s_Q + \boldsymbol{w}^H\boldsymbol{e} \end{aligned} \tag{20}$$

from where it can be seen that the condition of receiving the signal from one direction θ_m (known as the *direction of interest* (DOI)) as such, i.e. non-attenuated, implies the constraint

$$\boldsymbol{w}^H\boldsymbol{a}(\theta_m) = 1 \tag{21}$$

and for cancelling all other directions the conditions are:

$$\boldsymbol{w}^H\boldsymbol{a}(\theta) = 0 \qquad \forall \quad \theta \neq \theta_m. \tag{22}$$

In order to illustrate the motivation for the very constraints, we set the DOI equal to one of the source angles i.e. $\theta_m = \theta_1$ and look at (20) to see what would happen in this ideal situation:

$$y_F = 1s_1 + 0s_2 + \cdots + 0s_Q + \boldsymbol{w}^H\boldsymbol{e} = s_1 + \boldsymbol{w}^H\boldsymbol{e} \tag{23}$$

i.e. the output of the spatial filter equals to the power of the first source as such added with noise. If $\boldsymbol{e}$ is white noise any $\boldsymbol{w}$ will be as much correlated with it and thus the noise term $\boldsymbol{w}^H\boldsymbol{e}$ has equal power on average regardless of $\boldsymbol{w}$.

It is, however, not possible to design such $\boldsymbol{w}$ that would satisfy all constraints in (21) and (22). There are N variables in $\boldsymbol{w}$ which means that N constraints could, in principle, be satisfied. The equation (21) sets just a single constraint but (22) yields actually an infinite number of constraints. In fact (22) has to be true only for the source angles that do not correspond to the DOI but they are unknown at this point. The implementations of the spatial filtering idea such as Beamformer and MVDR choose to optimize a problem where the constraint (22) is replaced with another, computationally feasible constraint. This naturally leads to the value of $M(\theta) = \boldsymbol{w}^H\boldsymbol{a}(\theta)$ not being zero for all directions other than the DOI. $M(\theta)$ is known as the *response magnitude* [12, p. 277] and will be visualized later on.

As mentioned, the spatial spectrum is determined by the power of y_F. It can be expressed using the correlation matrix of the input signal $\boldsymbol{R} = \mathrm{E}[\boldsymbol{y}\boldsymbol{y}^H]$ as can be seen by starting from the definition of power:

$$\mathrm{E}[|y_F|^2] = \mathrm{E}[|\boldsymbol{w}^H\boldsymbol{y}|^2] = \mathrm{E}[\boldsymbol{w}^H\boldsymbol{y}(\boldsymbol{w}^H\boldsymbol{y})^*] = \mathrm{E}[\boldsymbol{w}^H\boldsymbol{y}\boldsymbol{y}^H\boldsymbol{w}] = \boldsymbol{w}^H\boldsymbol{R}\boldsymbol{w}. \tag{24}$$

When the covariance matrix is replaced with the estimate

$$\hat{\boldsymbol{R}} = \frac{1}{P}\sum_{p=1}^{P} \boldsymbol{y}(p)\boldsymbol{y}^H(p) \tag{25}$$

the spectrum estimate when $\boldsymbol{w}$ has been designed for θ_m yields

$$\mathcal{S}(\theta_m) = \boldsymbol{w}^H(\theta_m)\hat{\boldsymbol{R}}\boldsymbol{w}(\theta_m). \tag{26}$$

3.1.1 Beamformer

Beamformer is the most basic spatial spectrum estimation method. It is analogous to the periodogram in frequency spectrum estimation. The optimization problem behind Beamformer is the following:

$$\begin{aligned} &\text{minimize} \quad \boldsymbol{w}^H\boldsymbol{w} \\ &\text{subject to} \quad \boldsymbol{w}^H\boldsymbol{a}(\theta) = 1 \end{aligned} \tag{27}$$

where θ is the DOI. In this scheme $\boldsymbol{w}$ is not dependent on the data at all.

The solution yields

$$\boldsymbol{w} = \frac{\boldsymbol{a}(\theta)}{\boldsymbol{a}(\theta)^H\boldsymbol{a}(\theta)} = \frac{\boldsymbol{a}(\theta)}{N} \tag{28}$$

and placing this to equation (26) produces the spectral estimate

$$\mathcal{S}(\theta) = \frac{\boldsymbol{a}(\theta)^H\hat{\boldsymbol{R}}\boldsymbol{a}(\theta)}{N^2}. \tag{29}$$

3.1.2 Minimum Variance Distortionless Response (MVDR)

The *Capon*-method is also known as Minimun Variance Distortionless Response (MVDR). It was presented for frequency spectrum estimation by Capon in [1]. It is based on the idea that the filter $\boldsymbol{w}$ should be designed such that $E[|y_F|^2]$ (i.e. the overall power of the signal) is minimized subject to receiving a signal from one direction (angle θ) undistorted, leading to the minimization problem

$$\begin{aligned} &\text{minimize} \quad \mathrm{E}[|y_F|^2] \\ &\text{subject to} \quad \boldsymbol{w}^H\boldsymbol{a}(\theta) = 1. \end{aligned} \tag{30}$$

The minimization problem can be solved using Lagrange multipliers and the solution yields:

$$\boldsymbol{w} = \frac{\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)}{\boldsymbol{a}^H(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)} \tag{31}$$

and placing this to equation (26) produces the spectral estimate

$$\mathcal{S}(\theta) = \frac{1}{\boldsymbol{a}^H(\theta)\hat{\boldsymbol{R}}^{-1}\boldsymbol{a}(\theta)}. \tag{32}$$

MVDR is a superresolution method which means that its ability to resolve sources is better than that of Beamformer. This is simply due to the data dependency present in the optimization problem (30). The difference to Beamformer is visualized here with an example. The true wideband sources are at angles 50° and 58° and there is no additive noise. (Some noise is, however, introduced due to the use of DFT in the decomposition which is why the narrowband model is not exact.) We treat the subband corresponding to the center frequency 1312.5 Hz and use an ULA with 16 sensors separated by 0.5 meters and show the resulting spectra using the Beamformer and MVDR in Figure 6. It can be seen that the Beamformer is not able to resolve the sources in this case. This is due to the limited number of sensors and the design criterion of the Beamformer. Let us look closer at the situation using the idea of the response magnitude.

As described earlier, the response magnitude shows how a filter designed for a certain DOI behaves for other angles. Figure 7 illustrates the difference between Beamformer and MVDR. As the design criteria for Beamformer are not dependent on the data, the shape of the response magnitude curve is always the same. This can lead to poor close source separation ability. For example, in the center plot of Figure 7 we can see that when the DOI angle is 54° the response magnitude for the true sources is $|\boldsymbol{w}^H\boldsymbol{a}(\theta_q)| \approx 0.8$. When comparing this to (20) it can be concluded that a lot of source energy is captured by this DOI although we would not want any of it. This 80 % of the power of both sources taken into the spectrum is more than at either of the true source angles and is the reason why the Beamformer spectrum peaks at 50° in the example.

MVDR on the other hand is dependent on the data and is able to attenuate the true sources which are not desired to have an effect on the current DOI. The respective value for the center plot for MVDR is only about 0.2. The response magnitude is large for such angles that do not contain a source such as angles $> 58°$ on the rightmost plot. This is fine since only the noise power is captured in there.

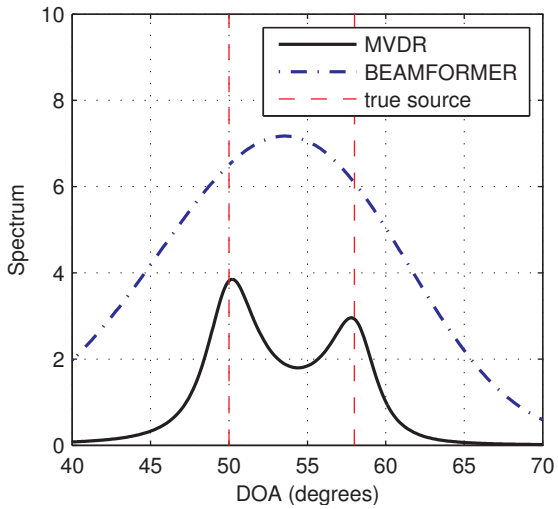


Figure 6: The spectra of Beamformer and MVDR. There are sources at 50° and 58°.

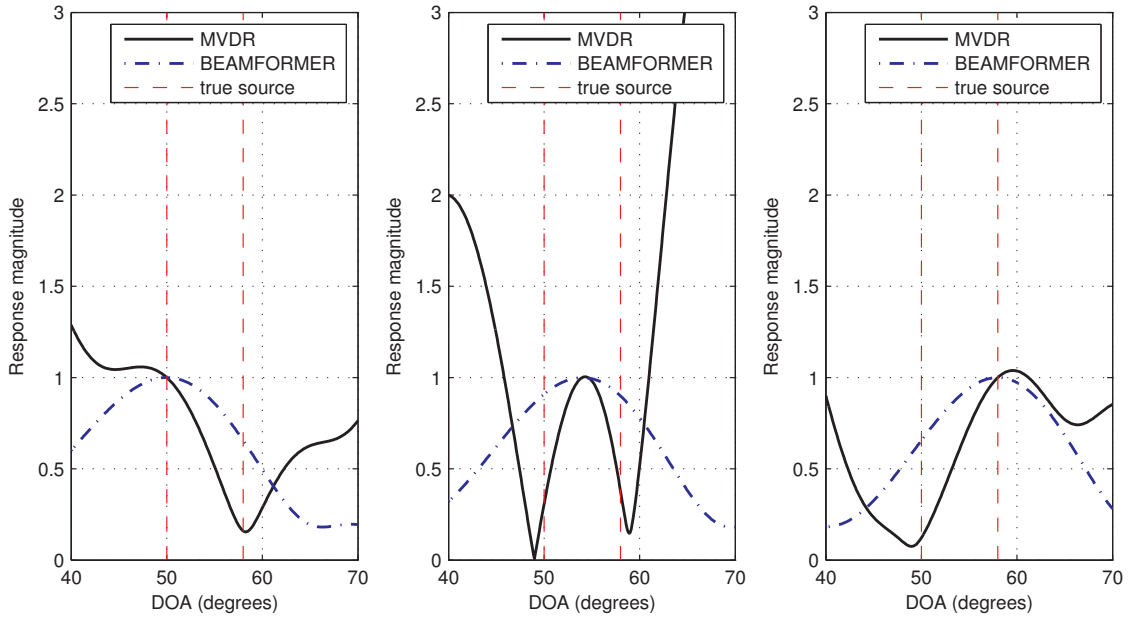


Figure 7: The response magnitudes of Beamformer and MVDR. There are sources at 50° and 58°. $\boldsymbol{w}$ is designed for 50° on the left plot, for 54° on the middle plot and for 58° on the right plot.

3.2 Subspace based methods

A different approach to the narrowband DOA estimation problem is the idea of subspace processing. By processing the covariance matrix of the measurements using the eigenvalue decomposition, the signal can be split into the noise and signal subspaces which in turn can be utilized to produce a pseudospectrum (i.e. a spectrum that does not give true power estimates) with a high resolution ability. Although there are other subspace methods such as Esprit, we choose to focus on MUSIC.

3.2.1 Multiple Signal Classification (MUSIC)

MUltiple SIgnal Cassification (MUSIC) first presented in [2] is considered to be a parametric method in frequency estimation because it is based on a parametric sinusoidal model. For DOA estimation this sinusoidal model corresponds to the model described by (17).

Let us break down the structure of the covariance matrix of the input signal by using the signal model (17):

$$\begin{aligned}
\boldsymbol{R} = \mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^H\right] &= \mathrm{E}\left[(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{e})(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{e})^H)\right] \\
&= \mathrm{E}\left[(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{e})(\boldsymbol{x}^H\boldsymbol{A}^H+\boldsymbol{e}^H)\right] \\
&= \mathrm{E}\left[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^H\boldsymbol{A}^H+\boldsymbol{A}\boldsymbol{x}\boldsymbol{e}^H+\boldsymbol{e}\boldsymbol{x}^H\boldsymbol{A}^H+\boldsymbol{e}\boldsymbol{e}^H\right] \\
&= \mathrm{E}[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^H\boldsymbol{A}^H]+\mathrm{E}[\boldsymbol{A}\boldsymbol{x}\boldsymbol{e}^H]+\mathrm{E}[\boldsymbol{e}\boldsymbol{x}^H\boldsymbol{A}^H]+\mathrm{E}[\boldsymbol{e}\boldsymbol{e}^H] \\
&= \boldsymbol{A}\,\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^H]\boldsymbol{A}^H+\boldsymbol{A}\,\mathrm{E}[\boldsymbol{x}\boldsymbol{e}^H]+\mathrm{E}[\boldsymbol{e}\boldsymbol{x}^H]\boldsymbol{A}^H+\mathrm{E}[\boldsymbol{e}\boldsymbol{e}^H]
\end{aligned}$$

Taking into account that noise and the source signals are not correlated and denoting the $Q \times Q$ source covariance matrix with $\boldsymbol{S}$ and noise covariance with $\boldsymbol{E}_c$

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H + \boldsymbol{E}_c. \tag{33}$$

If we further assume that the noise is white complex noise, $\boldsymbol{E}_c$ is simplified and the parametric form becomes

$$\boldsymbol{R}(\omega_k) = \boldsymbol{A}(\omega_k)\boldsymbol{S}(\omega_k)\boldsymbol{A}(\omega_k)^H + \sigma_{e,k}^2\boldsymbol{I} \tag{34}$$

where we have emphasized that for a possible wideband problem, this is true for a single frequency band. If the sources are uncorrelated, which is in practice true for our scenario, $\boldsymbol{S}(\omega_k)$ is a diagonal matrix. The crucial feature for MUSIC is that the rank of $\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H$ is equal to the number of sources Q which results simply from the dimensions of the matrices. Thus, $\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H$ has Q nonzero eigenvalues and $N-Q$ eigenvalues equal to zero. Since $\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H$ is also positive semidefinite the nonzero eigenvalues are positive. The noise term of (34) has N eigenvalues equal to the noise variance $\sigma_{e,k}^2$. $\boldsymbol{R}$ has the added eigenvalues of the two terms.

The eigenvalue decomposition (EVD) of $\boldsymbol{R}$

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}\boldsymbol{U}^{-1} \tag{35}$$

holds the eigenvectors in $\boldsymbol{U}$ and eigenvalues on the diagonal of $\boldsymbol{V}$.

$$\boldsymbol{V} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \lambda_N \end{bmatrix} \tag{36}$$

From now on, it is assumed that the eigenvalues are sorted in descending order i.e. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. $\boldsymbol{U}$ can be split into two parts: $\boldsymbol{U}_S$ has the Q first eigenvectors corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_Q$ and $\boldsymbol{U}_N$ holds the rest of the eigenvectors corresponding to the eigenvalues $\lambda_{Q+1}, \lambda_{Q+2}, \dots, \lambda_N$. $\boldsymbol{U}_S$ is known as the signal subspace and $\boldsymbol{U}_N$ as the noise subspace.

In order to derive the idea behind MUSIC, we start by the definition of eigenvalues:

$$\boldsymbol{R}\boldsymbol{U}_N = \boldsymbol{U}_N \begin{bmatrix} \lambda_{Q+1} & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \lambda_N \end{bmatrix} = \boldsymbol{U}_N \begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix} = \sigma_e^2 \boldsymbol{U}_N. \tag{37}$$

On the other hand, multiplying the form of $\boldsymbol{R}$ in (34) with $\boldsymbol{U}_N$ yields

$$\boldsymbol{R}\boldsymbol{U}_N = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H\boldsymbol{U}_N + \sigma_e^2\boldsymbol{U}_N. \tag{38}$$

From (37) it follows:

$$\boldsymbol{R}\boldsymbol{U}_N = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H\boldsymbol{U}_N + \boldsymbol{R}\boldsymbol{U}_N \Rightarrow \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H\boldsymbol{U}_N = 0 \tag{39}$$

Since $\boldsymbol{A}\boldsymbol{S}$ has full rank (the steering vectors are linearly independent)

$$\boldsymbol{A}^H\boldsymbol{U}_N = 0. \tag{40}$$

The true angles of the sources $\theta_1, \dots, \theta_Q$ then satisfy

$$\boldsymbol{a}(\theta_q)\boldsymbol{U}_N\boldsymbol{U}_N^H\boldsymbol{a}^H(\theta_q) = 0 \tag{41}$$

so the MUSIC pseudospectrum that peaks very sharply around the true angles when using the estimated covariance matrix to obtain $\boldsymbol{U}_N$ can be defined as the inverse of this

$$\mathcal{S}(\theta) = \frac{1}{\boldsymbol{a}(\theta)\boldsymbol{U}_N\boldsymbol{U}_N^H\boldsymbol{a}^H(\theta)}. \tag{42}$$

3.2.2 Selecting the number of sources

A peculiarity of MUSIC is that it requires a decision for the number of sources Q prior to performing the final spectrum estimation. This decision can have a major impact on the result. The information theoretic criterion based on Minimum

Description Length for detecting the number of sources was derived by Wax and Kailath in [13]. It assumes the white noise model for the covariance matrix of the noise and has the expression

$$\mathrm{MDL}(q) = -\log\left(\left(\frac{\prod_{d=q+1}^{N}\lambda_d^{1/(N-q)}}{\frac{1}{N-q}\sum_{d=q+1}^{N}\lambda_d}\right)^{(N-q)P}\right) + \frac{1}{2}q(2N-q)\log P. \tag{43}$$

The estimate for the number of sources can be selected as the minimum of the criterion i.e.

$$\hat{Q} = \underset{q=1,2,\ldots,N}{\arg\min}\,\mathrm{MDL}(q). \tag{44}$$

3.3 Sparse methods

The DOA estimation is inherently sparse: If a grid of candidate angles is constructed the true angles of the sources should be active and all the others inactive. Mathematically, the elements of the solution that correspond to the active angles are non-zero and the ones that correspond to the inactive angles are zero. This kind of a solution is called sparse. With the theory built around ℓ_1-optimization, sparsity has been a hot topic in the fields of signal processing and applied mathematics for the last decade.

For DOA estimation, Malioutov et al. proposed a method called ℓ_1-SVD in [7] which is based on the so called basis pursuit method for solving sparse systems. Another, more computationally efficient route for solving the sparse problems is the field of greedy algorithms. Such algorithms have also been applied to the DOA estimation problem and are covered in this section.

The signal model for a narrowband DOA problem is given in (18). In the equation, $\boldsymbol{Y}$ is known while $\boldsymbol{A}$, $\boldsymbol{X}$ and $\boldsymbol{E}$ are unknown. Something, however, is known about $\boldsymbol{A}$: It holds steering vectors as columns. The angles that define the steering vectors are unknown but we can construct a fat (i.e. one having more columns than rows) matrix $\boldsymbol{B}$ that consists of a number of candidate steering vectors. The resulting system of equations

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Z} + \boldsymbol{E} \tag{45}$$

is underdetermined – the number of solutions $\boldsymbol{Z}$ is infinite. Figure 9 illustrates the dimensions of the sparse system and Figure 8 shows correspondingly the dimensions of the model equation (18).

Let the grid of angles used to construct the dictionary be $\vartheta_1, \ldots, \vartheta_M$. Naturally, the

exact solution to (45) is then the one where

$$Z_{mp} = \begin{cases} X_{qp}, & \vartheta_m = \theta_q \\ 0, & \text{otherwise} \end{cases} \qquad q = 1, \ldots, Q \quad p = 1, \ldots, P. \tag{46}$$

This relation is depicted in Figure 9 by the colored columns which are the same steering vectors of which $\boldsymbol{A}$ in Figure 8 consists. Sure it is possible that the angle θ_q is not exactly on the grid. We, however, do not focus on this problem but assume that the grid is fine enough for the equality to hold well enough. Grid refinement techniques that attack this problem have been proposed in the literature.

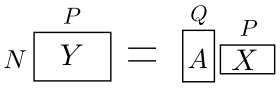


Figure 8: The dimensions of the model equation.

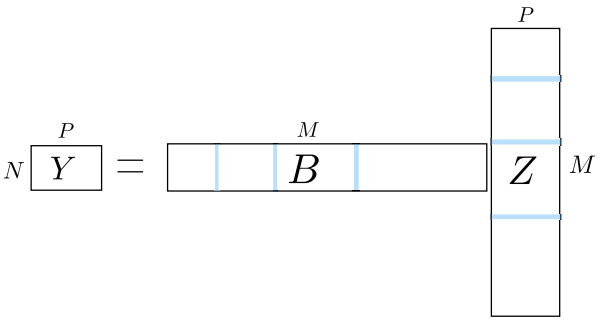


Figure 9: The dimensions of the sparse equation.

For simplicity, let us treat only one snapshot and denote the snapshot $\boldsymbol{y}$ and the sparse solution vector $\boldsymbol{z}$. If (45) was noiseless, the optimal (sparse) solution could be formulated as the following optimization problem:

$$\boldsymbol{z}_{\text{opt}} = \arg\min_{\boldsymbol{z}} \quad \|\boldsymbol{z}\|_{\ell_0} \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{B}\boldsymbol{z}. \tag{47}$$

The ℓ_0-norm in fact counts the number of nonzero entries which means that the solution is the sparsest possible. There are however no polynomial time algorithms to solve this problem so it needs to be relaxed. The methods for this are the so-called Basis Pursuit and the group of greedy algorithms.

3.4 ℓ_1-SVD

The work done by e.g. Emmanuel Candes and David Donoho relating to the field of *compressed sensing* proves that in sufficient conditions the ℓ_0-norm can be replaced with the more tractable ℓ_1-norm and the solution is still exact. When the presence of noise is also taken into account, the optimization problem yields

$$\boldsymbol{z}_{\text{opt}} = \arg\min_{\boldsymbol{z}} \quad \|\boldsymbol{z}\|_{\ell_1} \quad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{z}\|_2^2 \leq \hat{\sigma}_e^2 \tag{48}$$

which can also be expressed in the unconstrained form

$$\boldsymbol{z}_{\text{opt}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{z}\|_2^2 + \lambda\|\boldsymbol{z}\|_{\ell_1} \tag{49}$$

with λ being dependent on $\hat{\sigma}_e$.

The optimization problems encountered in compressed sensing in general follow the form of (48) where the measurements consist of a single vector. The DOA problem differs from this in the sense that the measurements consist of multiple snapshots instead. The snapshots should share the same sparsity (locations of nonzeros) and the model is thus called *jointly sparse*. Taking this into account with the ℓ_1 optimization problem, [7] proposes to solve

$$\boldsymbol{Z}_{\text{opt}} = \arg\min_{\boldsymbol{Z}} \|\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{Z}\|_F^2 + \lambda\|\boldsymbol{Z}^{(\ell_2)}\|_{\ell_1} \tag{50}$$

where the operation $\boldsymbol{Z}^{(\ell_2)}$ produces a column vector whose m-th element is equal to

$$\sqrt{\sum_{p=1}^{P} |Z_{mp}|^2}$$

i.e. the snapshots (rows) of $\boldsymbol{Z}$ are first combined using the ℓ_2-norm because there exists no sparsity in this direction. $\|\cdot\|_F$ denotes the Frobenius norm. $\boldsymbol{Z}$ is, in other words, row-sparse which is visible also in Figure 9. The optimization problem is convex and can be reformulated as a Second Order Cone Program (SOC) which can be solved with e.g. interior point methods.

Solving the problem requires a lot of computational power which is why there is a need to reduce the dimensionality. This is done in [7] via the singular value decomposition (SVD) by only keeping the signal subspace.

In the SVD factorization

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{V}^H \tag{51}$$

$\boldsymbol{U}$ holds the eigenvectors of $\boldsymbol{Y}\boldsymbol{Y}^H$ and $\boldsymbol{V}$ the eigenvectors of $\boldsymbol{Y}^H\boldsymbol{Y}$. $\boldsymbol{L}$ is a diagonal matrix where the elements are the square roots of the eigenvalues of both $\boldsymbol{Y}\boldsymbol{Y}^H$ and $\boldsymbol{Y}^H\boldsymbol{Y}$. The eigenvalues of the estimated covariance matrix (25) are thus given by

$$\lambda_n = \frac{1}{P} l_n^2 \tag{52}$$

where l_n is the n-th diagonal element of $\boldsymbol{L}$. These eigenvalues can now be used to select the order, $\hat{Q}$ with the MDL criterion (44) and estimate the power of noise

using the smallest eigenvalues:

$$\hat{\sigma}_e^2 = \frac{1}{N - \hat{Q}} \sum_{n=\hat{Q}+1}^{N} \lambda_n. \tag{53}$$

When the order, $\hat{Q}$ has been decided, it is possible to multiply both sides of (51) by $\boldsymbol{V}$ to get

$$\boldsymbol{YV} = \boldsymbol{UL} \tag{54}$$

where the right hand side can be reduced to only include the $\hat{Q}$ first eigenvectors to get the signal subspace i.e. $\boldsymbol{Y}_{\mathrm{SV}}$ is the $\hat{Q}$ first columns of $\boldsymbol{YV}$. Let $\boldsymbol{D}_Q = [\boldsymbol{I}_Q \boldsymbol{0}]$ be the matrix that contains concatenated identity matrix and a matrix of zeros. The new data matrix

$$\boldsymbol{Y}_{\mathrm{SV}} = \boldsymbol{YVD}_Q \tag{55}$$

still obeys the original model

$$\boldsymbol{Y}_{\mathrm{SV}} = \boldsymbol{BZ}_{\mathrm{SV}} + \boldsymbol{E}_{\mathrm{SV}} \tag{56}$$

where $\boldsymbol{Z}_{\mathrm{SV}} = \boldsymbol{ZVD}_Q$ and $\boldsymbol{E}_{\mathrm{SV}} = \boldsymbol{EVD}_Q$. The dimension of $\boldsymbol{Y}_{\mathrm{SV}}$ is reduced from $N \times P$ to $N \times \hat{Q}$. The value of P is typically 62 and the number of sources is typically just a few so the reduction is indeed significant. The complete ℓ_1-SVD problem becomes

$$\boldsymbol{Z}_{\mathrm{opt}} = \arg\min_{\boldsymbol{z}} \quad \|\boldsymbol{Z}_{\mathrm{SV}}^{(\ell_2)}\|_{\ell_1} \quad \text{subject to} \quad \|\boldsymbol{Y}_{\mathrm{SV}} - \boldsymbol{BZ}_{\mathrm{SV}}\|_F^2 \leq \beta^2 \tag{57}$$

which can also be expressed in the unconstrained form

$$\boldsymbol{Z}_{\mathrm{opt}} = \arg\min_{\boldsymbol{Z}_{\mathrm{SV}}} \|\boldsymbol{Y}_{\mathrm{SV}} - \boldsymbol{BZ}_{\mathrm{SV}}\|_F^2 + \lambda \|\boldsymbol{Z}_{\mathrm{SV}}^{(\ell_2)}\|_{\ell_1} \tag{58}$$

with λ being dependent on β.

A major issue considering the practical implementation of ℓ_1-SVD is the choice of β (or λ). In the form where no SVD is done, it would assumingly be natural to use just an estimate for the noise variance. SVD, however, complicates things in the sense that the matrix $\boldsymbol{V}$ is a function of $\boldsymbol{E}$. The ℓ_1-SVD article [7] proposes to assume a χ^2 distribution for the noise and then use an upper value of its confidence interval to set β. According to [7], this strategy only works for high SNR, though. It shows a curve of how $\|\boldsymbol{E}_{\mathrm{SV}}\|_F$ relates to σ_e out of which it can be concluded that when the sources are close and the SNR is near 0 dB, the relation is somewhat unpredictable while for high SNR the ratio $\frac{\|\boldsymbol{E}_{\mathrm{SV}}\|_F}{\sigma_e}$ is constant. This observation about the constant is used in our implementation of ℓ_1-SVD such that the estimate for $\|\boldsymbol{E}_{\mathrm{SV}}\|_F$ is set to $\beta = C\hat{\sigma}_e$ with the constant C being an input parameter for the

algorithm. The result of choosing β poorly can result according to [7] in a spectrum that has strong spurious peaks.

So how does ℓ_1-SVD compare to other narrowband estimation methods? It is obviously much more complicated to implement than MVDR or MUSIC because of the numerical optimization needed and in practice it is also much more computationally complex. However, if it could outperform MUSIC and MVDR resolutionwise it would be very interesting. The article [7] gives a single example where it can resolve sources closer than MUSIC but can not unfortunately give convincing statistical comparisons. The experiments conducted with ℓ_1-SVD were done as described by Algorithm 1. A simulated scenario where the true sources are moving close to each other is shown in Figure 10. The constant $C = 5$ (see Algorithm 1) was chosen manually to give a good result. In the figure, the peaks from the spectra given by MUSIC, MVDR and ℓ_1-SVD have been extracted. This kind of performance where ℓ_1-SVD is better than MVDR but can only almost be as good as MUSIC was found to be typical when C was selected well.

Input : The measurements $\boldsymbol{Y}$, angle grid $\vartheta_1, \ldots, \vartheta_M$, constant C.

1 Compute the SVD of $\boldsymbol{Y}$: $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{L}\boldsymbol{V}^H$.

2 Compute the eigenvalues such as in (52) and estimate the number of sources $\hat{Q}$ with the MDL criterion (44).

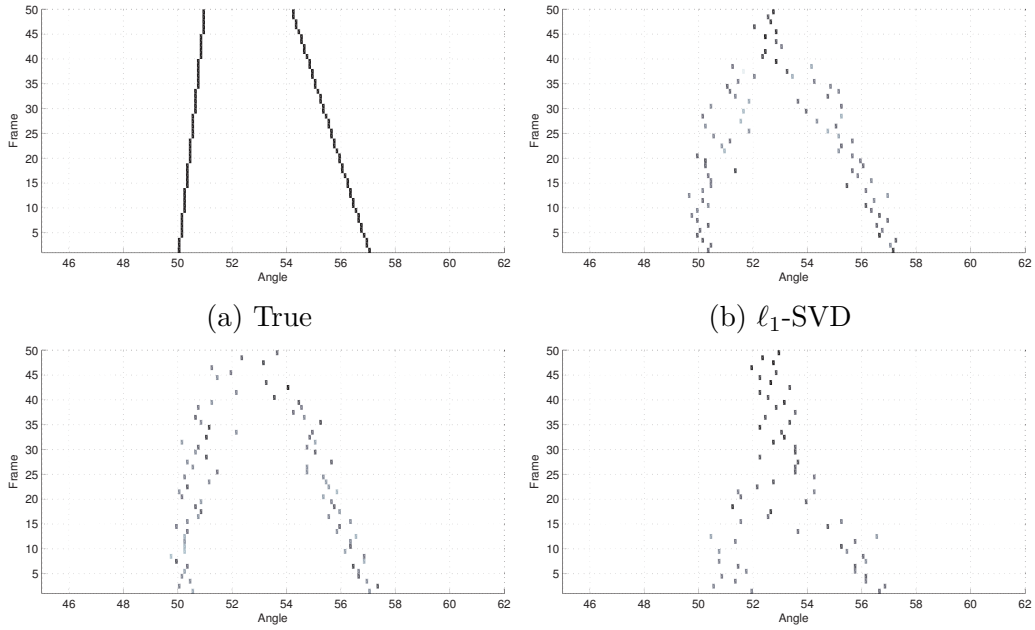
3 Estimate the noise variance $\hat{\sigma}_e^2$ as the average of the $N - \hat{Q}$ smallest eigenvalues with (53).

4 Compute Y_{SV} as in (55).

5 Set $\beta = C\hat{\sigma}_e$ and use SeDuMi to solve the optimization problem (57).

Output: The solution $\boldsymbol{Z}_{\mathrm{opt}}$ to (57) as the spectrum estimate $\mathcal{S}$.

Algorithm 1: ℓ_1-SVD



(a) True (b) ℓ_1-SVD

(c) MUSIC (d) MVDR

Figure 10: A narrowband comparison between ℓ_1-SVD, MUSIC and MVDR. The true sources are moving closer to each other with time (i.e. frame) as depicted in (a). SNR is 5 dB and the center frequency 1812.5 Hz. MVDR starts to see only a single peak in the middle of the two sources at about frame 23. MUSIC starts to fail at frame 40 and ℓ_1-SVD slightly earlier.

The optimization problem in the core of ℓ_1-SVD can be solved with a general numerical optimization package such as SeDuMi. A MATLAB implementation that uses SeDuMi for ℓ_1-SVD is provided by the authors of [9] at `http://code.soundsoftware.ac.uk/hg/doa-ad`. It has been originally written by Malioutov who is an author of [7]. This is the implementation used in this thesis to solve a narrowband DOA problem.

3.4.1 Greedy methods

Another well known strategy for solving the general sparse problem (47) is the family of greedy algorithms. The mostly widely known algorithm is called Orthogonal Matching Pursuit (OMP). The greedy nature of the algorithms means that they select one column from the dictionary iteratively by optimizing a local optimality criterion. Once a column is selected, it is held on. While greedy algorithms are simpler to understand and implement, they are still shown to have similar properties to ℓ_1 of exact recovery [14].

Greedy algorithms have been applied to the DOA problem in e.g. [9] and [10] and for the wideband case in [8]. Because the greedy algorithms are much less complex than ℓ_1-SVD, there is no need to use SVD to reduce dimensionality. They can operate

directly on the model (45). In the algorithms the snapshots are simply additively combined when choosing the best column, see e.g. (59).

3.4.2 Optimized Orthogonal Matching Pursuit

Optimized Orthogonal Matching Pursuit is a greedy algorithm for solving the sparse recovery problem. It was originally presented in the signal processing literature by Rebollo-Neira and Lowe in [15]. The same algorithm is sometimes also known as Orthogonal Least Squares (OLS, see [16]) which reflects its nature well. In OOMP, the set of selected columns is extended one at a time by such a column that has not been selected yet. The column that produces the smallest norm of the residual is selected. OOMP is presented in Algorithm 2.

Input : Measurements $\boldsymbol{y}(p)$, angle grid $\vartheta_1, \ldots, \vartheta_M$
Initialize $\Omega^0 = \emptyset$, $\boldsymbol{r}_1(p) = \boldsymbol{y}(p)$
for $q = 1, \ldots, N$ *or until a stopping criterion is met* **do**
  Define $\Omega_m^q = \Omega^{q-1} \cup m$
  $m_{\min} = \arg\min_{m \notin \Omega^{q-1}} \quad \sum_{p=1}^{P} \|\boldsymbol{r}_q(p) - \boldsymbol{A}_{\Omega_m^q} \boldsymbol{A}_{\Omega_m^q}^{\dagger} \boldsymbol{r}_q(p)\|^2$
  $\Omega^q = \Omega^{q-1} \cup m_{\min}$
  Update residual $\boldsymbol{r}_{q+1}(p) = \boldsymbol{y}(p) - \boldsymbol{A}_{\Omega^q} \boldsymbol{A}_{\Omega^q}^{\dagger} \boldsymbol{y}(p)$
end
Output: The set of selected angles Ω^q

Algorithm 2: OOMP

The similarity of OOMP to the Beamformer that was discussed in section 3.1.1 becomes clear when looking at the situation when no columns have been selected yet. For Optimized Orthogonal Matching Pursuit the criterion for selecting the index of the first column is

$$m = \arg\min_{m} \left\{ \min_{\boldsymbol{X}} \sum_{p=1}^{P} \|\boldsymbol{y}(p) - \boldsymbol{a}(\vartheta_m)\boldsymbol{x}(p)\|^2 \right\} \tag{59}$$

The inner optimization problem has the standard analytical LS-solution

$$\boldsymbol{x}(p) = (\boldsymbol{a}^H\boldsymbol{a})^{-1}\boldsymbol{a}^H\boldsymbol{y}(p) = \frac{1}{N}\boldsymbol{a}^H\boldsymbol{y}(p)$$

which yields

$$m = \arg\min_{m} \sum_{p=1}^{P} \|\boldsymbol{y}(p) - \frac{1}{N}\boldsymbol{a}(\vartheta_m)\boldsymbol{a}(\vartheta_m)^H\boldsymbol{y}(p)\|^2. \tag{60}$$

The norm can be expressed differently: (The angle ϑ_m and the snapshot index p are

omitted for simplicity since they stay constant.)

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{a}\boldsymbol{x}\|^2 &= (\boldsymbol{y} - \boldsymbol{a}\boldsymbol{x})^H(\boldsymbol{y} - \boldsymbol{a}\boldsymbol{x}) \\
&= (\boldsymbol{y} - \frac{\boldsymbol{a}\boldsymbol{a}^H}{N}\boldsymbol{y})^H(\boldsymbol{y} - \frac{\boldsymbol{a}\boldsymbol{a}^H}{N}\boldsymbol{y}) \\
&= (\boldsymbol{y}^H - \boldsymbol{y}^H\frac{\boldsymbol{a}\boldsymbol{a}^H}{N})(\boldsymbol{y} - \frac{\boldsymbol{a}\boldsymbol{a}^H}{N}\boldsymbol{y}) \\
&= \boldsymbol{y}^H\boldsymbol{y} - \frac{1}{N}\boldsymbol{y}^H\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{y} - \frac{1}{N}\boldsymbol{y}^H\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{y} + \boldsymbol{y}^H\frac{\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{a}\boldsymbol{a}^H}{NN}\boldsymbol{y} \\
&= \|\boldsymbol{y}\|^2 - \frac{1}{N}\boldsymbol{y}^H\boldsymbol{a}\boldsymbol{a}^H\boldsymbol{y} \\
&= \|\boldsymbol{y}\|^2 - \frac{1}{N}\boldsymbol{a}^H\boldsymbol{y}\boldsymbol{y}^H\boldsymbol{a}.
\end{aligned}$$

This result put into (60)

$$m = \arg\min_m \sum_{p=1}^{P} \left[\|\boldsymbol{y}(p)\|^2 - \frac{1}{N}\boldsymbol{a}(\vartheta_m)^H\boldsymbol{y}(p)\boldsymbol{y}(p)^H\boldsymbol{a}(\vartheta_m)\right] \tag{61}$$

can be expressed as

$$m = \arg\min_m \sum_{p=1}^{P} \|\boldsymbol{y}(p)\|^2 - \frac{P}{N}\boldsymbol{a}(\vartheta_m)^H\hat{\boldsymbol{R}}\boldsymbol{a}(\vartheta_m) \tag{62}$$

The first term does not depend on the angle ϑ_m so the optimization is equivalent to maximizing the latter term. This is in fact the same as the Beamformer spectrum (29) multiplied by $\frac{N}{P}$. The very same observation is made on [12, p. 288] in the context of Nonlinear Least Squares.

The first iteration of OOMP is similar to the Beamformer and can not be recovered from due to the greedy nature of the algorithm. OOMP can, however, use the information of the previously selected angles in the following iterations which can lead to better performance compared to the Beamformer.

3.4.3 Orthogonal Matching Pursuit

The best known greedy algorithm to solve the sparse recovery problem in general is known as Orthogonal Maching Pursuit. More information about it can be found in [17]. It relaxes the criterion of OOMP by only looking at the scalar product of the candidate column and the current residual. Geometrically this translates to selecting the column that has the smallest angle to the residual. OOMP, on the other hand, geometrically projects the candidate columns to the surface spanned by the orthogonal complement of the current signal estimate. OMP is presented in

Algorithm 3.

Input : Measurements $\boldsymbol{y}(p)$, angle grid $\vartheta_1, \ldots, \vartheta_M$

Initialize $\Omega^0 = \emptyset$, $\boldsymbol{r}_1(p) = \boldsymbol{y}(p)$

for $q = 1, \ldots, N$ *or until a stopping criterion is met* **do**

  $m_{\min} = \arg\min_m \quad \sum_{p=1}^{P} \frac{|\boldsymbol{a}(\vartheta_m)^H \boldsymbol{r}_q(p)|}{\|\boldsymbol{a}(\vartheta_m)\|} \qquad \forall \quad m \notin \Omega^{q-1}$

  $\Omega^q = \Omega^{q-1} \cup m_{\min}$

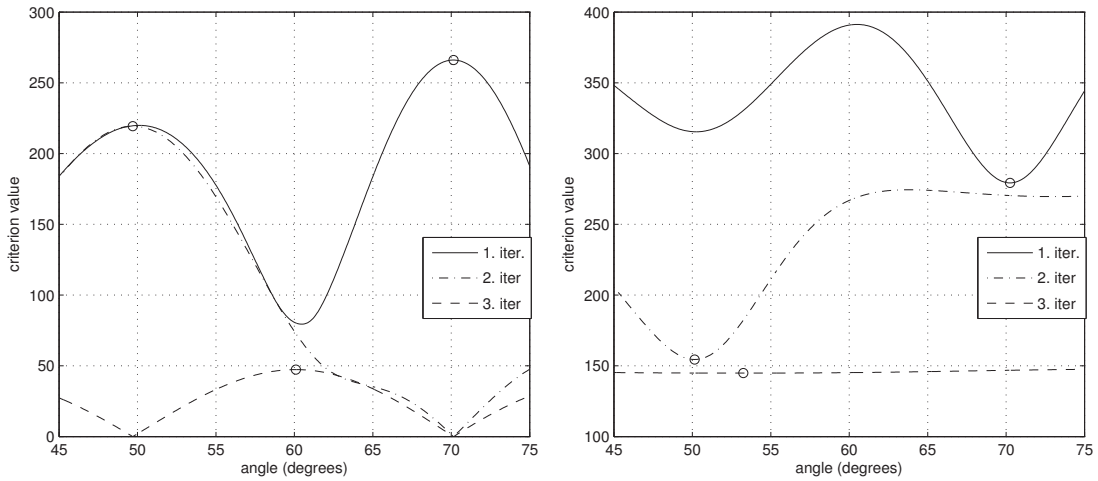
  Update residual $\boldsymbol{r}_{q+1}(p) = \boldsymbol{y}(p) - \boldsymbol{A}_{\Omega^q} \boldsymbol{A}_{\Omega^q}^{\dagger} \boldsymbol{y}(p)$

end

Output: The set of selected angles Ω^q

Algorithm 3: OMP

Also for OMP the observation made in Section 3.4.2 about the similarity of the first iteration criterion and the Beamformer criterion holds. The practical difference between the criteria and the iterative nature of OMP and OOMP can be seen in Figure 11. From it, it can be seen that if the iterations are continued too far, the selected nonzero location is somewhat arbitrary. An information theoretic criterion should be used to decide when to stop the iterations. From the criterion of OOMP at the third iteration, for example, it can be said that it is very flat and none of the locations offer a big change to the criterion which should be easy to detect.



(a) OMP (b) OOMP

Figure 11: An example of the difference between OMP and OOMP at different iterations of the algorithm. The two sources are at 50° and 70°. Note that for OMP the criterion is maximized and for OOMP minimized. The locations of the chosen positions at each iteration is noted with a circle.

3.4.4 Other sparse models

A narrowband DOA estimation method is presented in [18]. Instead of applying a sparse recovery algorithm to the data model (18), it uses directly the covariance

equation (34). The unknown matrix $\boldsymbol{Z}$ is in there defined to be the $\boldsymbol{S}\boldsymbol{A}^H$ part of the covariance equation. We experimented with the method and do not see much difference to the model used by OMP and OOMP.

Another method for sparse overcomplete representation for DOA estimation was also proposed in [19]. It however makes the assumption that the unified correlation functions of the incident signals are identical which is hardly the case in our scenario.

4 Wideband methods

In this section we show algorithms for the full wideband problem which is described by equation (17). The algorithms produce a wideband spectrum estimate $\mathcal{S}$ which is evaluated at the grid points $\vartheta_1, \vartheta_2, \ldots, \vartheta_M$. Typically only a subset of the subbands $\omega_1, \omega_2, \ldots, \omega_K$ is used. This is because the low frequencies often have very poor resolution and the high frequencies are prone to become aliased. The set of used subband indices is denoted by $\mathcal{K}$.

In the next subsections, the most commonly used wideband methods found in the signal processing literature are described. They can be divided into three categories: incoherent, coherent and sparse methods. The incoherent methods are most straightforward ones. They apply a narrowband method separately for each subband and then use a strategy for combining these independent estimates. The coherent methods are based on the idea of cohering the frequencies into a single frequency for which a narrowband-like estimation method can then be applied. The sparse methods are extensions to the narrowband sparse methods described in Section 3.3.

4.1 Incoherent wideband methods

In incoherent methods, each subband ω_k is processed individually to acquire the narrowband spectrum estimate $\mathcal{S}_k$. The subband spectra contain the power of the signal into a direction if a spatial filtering method (discussed in Section 3.1) is used to acquire them. If we take into account the power conserving property of the DFT (14) used in decomposing the signal it is possible to say that the power estimate of the wideband source signal is simply the summation of the subband spectra.

We use the method of combining the subbands additively for MVDR and for MUSIC even though the latter is not based on the idea of spatial filtering and gives pseudospectra instead of true power estimates. The method is sometimes known in the literature as arithmetic averaging or e.g. incoherent-MUSIC.

4.1.1 Summation based MVDR

The algorithm describing the summation based MVDR (MVDR-S) is given in Algorithm 4.

Input : The measurements $\boldsymbol{Y}_k$, angle grid $\vartheta_1, \ldots, \vartheta_M$.
for $k \in \mathcal{K}$ **do**
  Estimate the covariance matrix $\hat{\boldsymbol{R}}_k = \frac{1}{P}\boldsymbol{Y}_k\boldsymbol{Y}_k^H$.
  for $m = 1, \ldots, M$ **do**
    $\mathcal{S}_k(\vartheta_m) = \frac{1}{\boldsymbol{a}^H(\vartheta_m)\hat{\boldsymbol{R}}_k^{-1}\boldsymbol{a}(\vartheta_m)}$
  end
end
for $m = 1, \ldots, M$ **do**
  $\mathcal{S}(\vartheta_m) = \sum_{k\in\mathcal{K}} \mathcal{S}_k(\vartheta_m)$
end
Output: $\mathcal{S}$

Algorithm 4: MVDR-S

4.1.2 Summation based MUSIC

The algorithm describing the summation based MUSIC (MUSIC-S) is given in Algorithm 5.

Input : The measurements $\boldsymbol{Y}_k$, angle grid $\vartheta_1, \ldots, \vartheta_M$.
for $k \in \mathcal{K}$ **do**
  Estimate the covariance matrix $\hat{\boldsymbol{R}}_k = \frac{1}{P}\boldsymbol{Y}_k\boldsymbol{Y}_k^H$.
  Compute the EVD of $\hat{\boldsymbol{R}}(\omega_k) = \hat{\boldsymbol{U}}(\omega_k)\hat{\boldsymbol{V}}(\omega_k)\hat{\boldsymbol{U}}^{-1}(\omega_k)$.
  Estimate the number of sources $\hat{Q}_k$ using (44).
  Take the eigenvectors in $\hat{\boldsymbol{U}}(\omega_k)$ that correspond to the $N - \hat{Q}_k$ smallest eigenvalues and denote the matrix containing them with $\hat{\boldsymbol{U}}_N(\omega_k)$.
  for $m = 1, \ldots, M$ **do**
    $\mathcal{S}_k(\vartheta_m) = \frac{1}{\boldsymbol{a}^H(\vartheta_m,\omega_k)\hat{\boldsymbol{U}}_N(\omega_k)\hat{\boldsymbol{U}}_N^H(\omega_k)\boldsymbol{a}(\vartheta_m,\omega_k)}$.
  end
end
for $m = 1, \ldots, M$ **do**
  $\mathcal{S}(\vartheta_m) = \sum_{k\in\mathcal{K}} \mathcal{S}_k(\vartheta_m)$
end
Output: $\mathcal{S}$

Algorithm 5: MUSIC-S

4.2 Coherent wideband methods

The idea about coherent wideband DOA estimation was presented first by Wang and Kaveh in 1988. The algorithm called *Coherent signal-subspace processing* (CSSM) was introduced in [20]. Instead of combining the individual DOA estimates from

different subbands (as is done in incoherent methods) the coherent method performs the DOA estimation based on a single covariance matrix which is a representative of the wideband signal. This matrix is sometimes referred to as universal spatial covariance matrix (USCM). The goal is to form a model similar to (34) but for a single reference frequency ω_0 i.e.

$$\boldsymbol{R}(\omega_0) = \boldsymbol{A}(\omega_0)\boldsymbol{S}(\omega_0)\boldsymbol{A}(\omega_0)^H + \sigma_{e,0}^2\boldsymbol{I} \tag{63}$$

and to apply a subspace method such as MUSIC for this.

The basic assumption for the coherent methods is that a steering matrix can be transformed into another steering matrix with reference frequency ω_0 with an $N \times N$ transformation matrix $\boldsymbol{T}(\omega_k)$

$$\boldsymbol{T}(\omega_k)\boldsymbol{A}(\omega_k) = \boldsymbol{A}(\omega_0). \tag{64}$$

The derivation of the model begins by defining the transformed measurements

$$\boldsymbol{Y}_t(\omega_k) = \boldsymbol{T}(\omega_k)\boldsymbol{Y}(\omega_k). \tag{65}$$

Taking the covariance of the transformed measurements and noting that the expectation does not depend on the transformation matrix we arrive at

$$\sum_{k\in\mathcal{K}} \mathrm{E}[\boldsymbol{Y}_t(\omega_k)\boldsymbol{Y}_t(\omega_k)^H] = \sum_{k\in\mathcal{K}} \mathrm{E}[\boldsymbol{T}(\omega_k)\boldsymbol{Y}(\omega_k)\boldsymbol{Y}(\omega_k)^H\boldsymbol{T}(\omega_k)^H] = \sum_{k\in\mathcal{K}} \boldsymbol{T}(\omega_k)\boldsymbol{R}(\omega_k)\boldsymbol{T}(\omega_k)^H.$$

Substituting the parametric model for $\boldsymbol{R}(\omega_k)$ from (34)

$$\sum_{k\in\mathcal{K}} \boldsymbol{T}(\omega_k)\boldsymbol{R}(\omega_k)\boldsymbol{T}(\omega_k)^H = \sum_{k\in\mathcal{K}} \left(\boldsymbol{T}(\omega_k)\boldsymbol{A}(\omega_k)\boldsymbol{S}(\omega_k)(\boldsymbol{T}(\omega_k)\boldsymbol{A}(\omega_k))^H + \sigma_k^2\boldsymbol{T}(\omega_k)\boldsymbol{T}(\omega_k)^H\right)$$

and utilizing the assumption (64)

$$\sum_{k\in\mathcal{K}} \boldsymbol{T}(\omega_k)\boldsymbol{R}(\omega_k)\boldsymbol{T}(\omega_k)^H = \sum_{k\in\mathcal{K}} \left(\boldsymbol{A}(\omega_0)\boldsymbol{S}(\omega_k)\boldsymbol{A}(\omega_0)^H + \sigma_{e,k}^2\boldsymbol{T}(\omega_k)\boldsymbol{T}(\omega_k)^H\right)$$

The transform matrices used in this thesis are unitary which further simplifies the model

$$\sum_{k\in\mathcal{K}} \boldsymbol{T}(\omega_k)\boldsymbol{R}(\omega_k)\boldsymbol{T}(\omega_k)^H = \boldsymbol{A}(\omega_0)\sum_{k\in\mathcal{K}} (\boldsymbol{S}(\omega_k))\boldsymbol{A}(\omega_0)^H + K\sum_{k\in\mathcal{K}} (\sigma_{e,k}^2)\boldsymbol{I} \tag{66}$$

If we denote

$$\boldsymbol{S}(\omega_0) = \sum_{k\in\mathcal{K}} \boldsymbol{S}(\omega_k),$$

$$\sigma_{e,0}^2 = K \sum_{k \in \mathcal{K}} \sigma_{e,k}^2$$

and

$$\boldsymbol{R}(\omega_0) = \sum_{k \in \mathcal{K}} \boldsymbol{T}(\omega_k) \boldsymbol{R}(\omega_k) \boldsymbol{T}(\omega_k)^H \tag{67}$$

it can be noticed that (66) is exactly (63) which was the goal. Now, a subspace method such as MUSIC is applicable to the estimate

$$\hat{\boldsymbol{R}}(\omega_0) = \sum_{k \in \mathcal{K}} \boldsymbol{T}(\omega_k) \hat{\boldsymbol{R}}(\omega_k) \boldsymbol{T}(\omega_k)^H \tag{68}$$

and by such a wideband spectrum can be acquired.

In the coherent subspace processing, the challenge is to design the transform matrices. In practice, the angles defining the steering vectors that form both matrices $\boldsymbol{A}(\omega_0)$, $\boldsymbol{A}(\omega_k)$ in equation (64) are unknown (they are exactly what is being estimated, in fact!). Like in other methods, we can build steering matrices that contain candidate steering vectors. Let such a candidate steering matrix be denoted by $\boldsymbol{B}(\omega_k)$. If $\boldsymbol{B}$ is of size $N \times N$, then $\boldsymbol{T}(\omega_k) = \boldsymbol{B}(\omega_0)\boldsymbol{B}(\omega_k)^{-1}$. In [20] ideas for 'fulfilling' the candidate matrix $\boldsymbol{B}$ by using initial estimates and other additional angles are given. Then, (64) holds approximately in the neighbourhood of these angles.

The simplest way to design a transform matrix is for (64) to hold exactly for a single angle θ. The resulting transform matrix is diagonal and the n-th diagonal entry is

$$a_n(\omega_0, \theta_m)/a_n(\omega_k, \theta_m) \tag{69}$$

where θ_m is the selected angle. This is the method proposed in [20] and [4] when all the source angles are concentrated within a beamwidth. The sweeping angle should be used only in this beamwidth restricted area. Beamwidth refers to the resolution limit for the Beamformer and is discussed in more detail in Section 7.1.

The performance of some transform matrices are discussed in [4]. The rotational signal subsace (RSS) focusing matrices proposed there have shown to be effective. They are used as a comparison in [6] and they outperform the proposed TOPS method for low SNR scenario which is true in the underwater surveillance application.

Instead of fulfilling the candidate matrices to be $N \times N$, it is proposed to solve

$$\begin{aligned} &\text{minimize} && \|\boldsymbol{B}(\omega_0, \boldsymbol{\alpha}) - \boldsymbol{T}(\omega_k)\boldsymbol{B}(\omega_j, \boldsymbol{\alpha})\|_F \\ &\text{subject to} && \boldsymbol{T}^H(\omega_k)\boldsymbol{T}(\omega_k) = \boldsymbol{I} \end{aligned} \tag{70}$$

where $\boldsymbol{\alpha}$ is a set of angles based on a pre-estimation that define the steering vectors

contained in $\boldsymbol{B}$. The RSS idea suggests that a solution for (70) is given by

$$\boldsymbol{T}(\omega_k) = \boldsymbol{V}(\omega_k)\boldsymbol{U}^H(\omega_k) \tag{71}$$

where $\boldsymbol{V}$ and $\boldsymbol{U}$ are the left and right singular vectors of $\boldsymbol{A}(\omega_k, \boldsymbol{\alpha})\boldsymbol{A}^H(\omega_0, \boldsymbol{\alpha})$. $\boldsymbol{\alpha}$ should contain angles $\{\theta_q, \theta_q - 0.25\text{BW}, \theta_q + 0.25\text{BW}\}$ where BW denotes beamwidth for each initial angle estimate θ_q. It is important to note that [4] claims that if there is only one initial angle estimate, the diagonal focusing matrix corresponding to (69) should be used!

Algorithm 6 summarizes the implementation of CSSM. As will be discussed in Section 7.1, the beamwidth is not unique for the wideband problem which is why the average of the used frequencies is used to evaluate the beamwidth. The formula can be compared to (119).

Input : The data $\boldsymbol{Y}_k$, the angle grid $\vartheta_1, \ldots, \vartheta_M$, the set of selected subbands $\mathcal{K}$.
Acquire initial estimates for θ_q and $\hat{Q}$ using MVDR-S.
Denoting $K = |\mathcal{K}|$, evaluate beamwidth BW $= \frac{c}{x_N 1/K \sum_{k\in\mathcal{K}} \omega_k}$.
Initialize $\hat{\boldsymbol{R}}(\omega_0)$ to zeros.
for $k \in \mathcal{K}$ **do**
    Estimate the covariance matrix $\hat{\boldsymbol{R}}(\omega_k)$ (25).
    if $\hat{Q} = 1$ **then**
        Build the diagonal focusing matrix $\boldsymbol{T}(\omega_k)$ like in (69) around θ_1.
    end
    else
        Build the RSS focusing matrix $\boldsymbol{T}(\omega_k)$ using (71) with BW as the beamwidth.
    end
    Update $\hat{\boldsymbol{R}}(\omega_0) = \hat{\boldsymbol{R}}(\omega_0) + \boldsymbol{T}(\omega_k)\hat{\boldsymbol{R}}(\omega_k)\boldsymbol{T}(\omega_k)^H$.
end
Estimate the number of sources using the AIC criterion (73).
Use MUSIC to estimate the wideband spectrum $\mathcal{S}$ using $\hat{\boldsymbol{R}}(\omega_0)$.
Output: $\mathcal{S}$

Algorithm 6: CSSM

4.3 The number of sources for the coherent methods

An information theoretic criterion similar to the one described in Section 3.2.2 can also be used to choose an estimate for the number of sources when the eigenvalues are computed from the universal spatial covariance matrix. This criterion, which is given in [20], is called the Akaike Information Criterion and suggests that the

estimate for the number of sources $\hat{Q}$ can be selected as the minimum of

$$\mathrm{AIC}(q) = (N-q)P \log \left(\frac{\frac{1}{N-q}\sum_{d=q+1}^{N} \lambda_d}{\left(\prod_{d=q+1}^{N} \lambda_d\right)^{1/(N-q)}} \right) + q(2N-q) \tag{72}$$

i.e.

$$\hat{Q} = \underset{q=1,2,\ldots,N}{\arg\min}\ \mathrm{AIC}(q). \tag{73}$$

4.3.1 WAVES algorithm

In 2001, a method called *Weighted Average of Signal Subspaces* (WAVES) was proposed by DiClaudio and Parisi in [5]. The paper lists a number of drawbacks for CSSM such as the sensitivity to modeling errors and not being able to focus on the entire field of view. They develop a statistic called *pseudodata matrix* referring to the framework of *robust statistics*.

They name (63) USCM, Universal Spatial Covariance Matrix, and propose to replace it with the pseudodata matrix

$$\hat{\boldsymbol{Z}} = Q^{-1/2} \left[\hat{\boldsymbol{Z}}_1, \ldots, \hat{\boldsymbol{Z}}_K \right] \tag{74}$$

where

$$\hat{\boldsymbol{Z}}_k = \boldsymbol{T}(\omega_k) \hat{\boldsymbol{U}}_s(\omega_k) \boldsymbol{P}(\omega_k) \tag{75}$$

$\hat{\boldsymbol{U}}_s(\omega_k)$ being the estimated signal subspace at frequency ω_k and $\boldsymbol{P}(\omega_k)$ a diagonal matrix with the diagonal elements equal to

$$P(\omega_k)_{qq} = \frac{\lambda_q(\omega_k) - \sigma_e^2}{\sqrt{\lambda_q(\omega_k)\sigma_e^2}}. \tag{76}$$

where $\lambda_q(\omega_k)$ is the q-th largest eigenvalue corresponding to the the q-th source at frequency ω_k. The effective rank of $\hat{\boldsymbol{Z}}$ should still be equal to the number of sources. Focusing errors and finite sample size makes it full rank. The reduced size SVD can be applied to it and Q principal left singular vectors represent the signal subspace, the rest $N - Q$ representing the noise subspace. The standard MUSIC algorithm can be applied to the noise subspace.

The implementation of the WAVES algorithm is summarized in Algorithm 7. Compared to CSSM, it is much more complex since it requires the compuation of the eigenvalue decomposition for each subband. The main difference to CSSM seems to be that only the signal subspace is used in each subband instead of the whole covariance matrix. Intuitively this should improve the estimation result when some sources are only active on some, but not all, of the subbands.

Input : The data $\boldsymbol{Y}_k$.
for $k \in \mathcal{K}$ **do**
  Estimate the covariance matrix (25).
  Compute the EVD of $\hat{\boldsymbol{R}}(\omega_k)$.
  Estimate the number of sources using the MDL criterion (44).
  Compute and save $\hat{\boldsymbol{Z}}_k$ as in (75) using the RSS transformation matrix (71).
end
Build $\hat{\boldsymbol{Z}}$ (74) using the saved submatrices.
Apply R-SVD to $\hat{\boldsymbol{Z}}$ using the maximum $\hat{Q}$ of the subbands as the estimate for the number of sources.
Use MUSIC to estimate the DOAs.
Output: $\mathcal{S}$

Algorithm 7: WAVES

4.3.2 TOPS algorithm

In 2006, a new algorithm called *Test Of Orthogonality of Projected Subspaces* (TOPS) was introduced in [6]. It can also be considered a coherent method since it uses a transformation matrix to cohere the frequencies. The transformation matrix is the diagonal type described by (69) and it is applied to each DOI separately. This is different than in CSSM or WAVES where the transformation matrix is only applied once. A similar strategy is also used in the Steered Covariance Method [21].

TOPS starts by choosing the reference frequency and the corresponding estimated signal subspace $\hat{\boldsymbol{U}}_s(\omega_0)$. It is mentioned in [6] that all the sources should overlap at least in one subband. This subband should assumably be selected for the reference frequency but there is no mention of how to choose this subband.

TOPS says that the range space of this signal subspace multiplied with the transformation matrix is equal to the range space of the true steering matrix of the k-th subband.

$$\mathcal{R}\left\{\boldsymbol{T}(\omega_k)\hat{\boldsymbol{U}}_s(\omega_0)\right\} = \mathcal{R}\left\{\boldsymbol{A}(\omega_k, \boldsymbol{\theta})\right\} \tag{77}$$

The core of the algorithm is to build a matrix

$$\boldsymbol{D} = \left[\boldsymbol{W}(\omega_1)^H \boldsymbol{U}_N(\omega_1) \mid \cdots \mid \boldsymbol{W}(\omega_K)^H \boldsymbol{U}_N(\omega_K)\right] \tag{78}$$

where

$$\boldsymbol{W}(\omega_k) = \boldsymbol{P}(\omega_k)\boldsymbol{T}(\omega_k)\boldsymbol{U}_S(\omega_0). \tag{79}$$

$\boldsymbol{D}$ should be rank-deficient when the transformation matrix is designed to one of the true angles. The part *projected* of the name comes from the use of a projection

matrix $\boldsymbol{P}(\omega_k)$. For the DOI θ_m it is

$$\boldsymbol{P}(\omega_k) = \boldsymbol{I} - \frac{\boldsymbol{a}(\omega_k, \theta_m)\boldsymbol{a}^H(\omega_k, \theta_m)}{N} \tag{80}$$

i.e. the projection onto the null space of $\boldsymbol{a}$. It is mentioned that this reduces some error terms in the TOPS matrix $\boldsymbol{D}$.

In practice, the TOPS matrix never becomes exactly rank deficient which is why the measure of rank-deficiency is defined as the smallest singular value of $\boldsymbol{D}$. The TOPS pseudospectrum is then the inverse of this rank-deficiency measure

$$\mathcal{S}(\theta_m) = \frac{1}{\sigma_{\min}(\theta_m)}. \tag{81}$$

The introduction of TOPS is argued in [6] mostly by the fact that other coherent methods (CSSM, WAVES) are unable to find an exact solution to the wideband problem when there is no noise. This is because the cohering causes noise by itself. TOPS, however, is able to overcome this. The result section of [6] compares TOPS to CSSM and WAVES and is only able to overperform them in high SNR scenario. This fact and the difficulty of selecting the correct reference frequency does not promise great results in our problem.

Algorithm 8 summarizes the TOPS algorithm. TOPS is most complex of the coherent methods since the transformation matrices have to be evaluated for each DOI.

Input : The data $\boldsymbol{Y}_k$, the angle grid $\vartheta_1, \vartheta_2, \ldots, \vartheta_M$, the reference frequency ω_0, the set of used subbands $\mathcal{K}$.

1 Compute the signal subspace $\hat{\boldsymbol{U}}_S(\omega_0)$ and noise subspace $\hat{\boldsymbol{U}}_N(\omega_k)$ for $k = 1, \ldots, K$ from the sample covariance of $\boldsymbol{Y}_k$.
2 **for** $m = 1, 2, \ldots, M$ **do**
3 Generate $\boldsymbol{D}$ for ϑ_m using (78), (79) and (80).
4 Compute the singular values of $\boldsymbol{D}$.
5 Estimate the pseudospectrum using the smallest singular value: $\mathcal{S}(\vartheta_m) = \frac{1}{\sigma_{\min}(\vartheta_m)}$.
6 **end**

Output: $\mathcal{S}$

Algorithm 8: TOPS

4.4 Sparse wideband methods

Solving the wideband problem with the sparse methods presented in Section 3.3 is dealt in this section. First, ℓ_1-SVD is discussed and then practical implementations for the greedy methods are given.

4.4.1 On the applicability of ℓ_1-SVD

The ℓ_1-SVD algorithm described in Section 3.4 is not easy to extend to the wideband problem. The original paper [7] mentions the problem but does not provide a method for combining the subband estimates. It shows a result of processing the subbands individually but this is not sufficient as such. In his master thesis [22], Malioutov provides the idea of jointly processing the wideband problem by stacking the measurements

$$\boldsymbol{Y}_{\mathrm{SV}} = \begin{bmatrix} \boldsymbol{Y}_{\mathrm{SV}}(\omega_1)^T & \boldsymbol{Y}_{\mathrm{SV}}(\omega_2)^T & \dots & \boldsymbol{Y}_{\mathrm{SV}}(\omega_K)^T \end{bmatrix}^T \tag{82}$$

and the matrix

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}(\omega_1) & & \\ & \ddots & \\ & & \boldsymbol{B}(\omega_K) \end{bmatrix} \tag{83}$$

and then solving the optimization problem similar to the narrowband case (57). Like [22] mentions, as is, this leads to sparsity also in the frequency direction. According to the same source, this can be avoided by combining the frequencies corresponding to a DOI by ℓ_2-norm in the same fashion that was done for the snapshots in the narrowband ℓ_1-SVD. By stacking the matrices as in (82) and (83), the size of the problem for our scenario becomes infeasible. Typically the size of the matrix $\boldsymbol{B}$ would be $NK \times MK = 16 \cdot 40 \times 180 \cdot 40 = 640 \times 7200$ which is too much to be solved online (as is discussed in Section 7.3) at least with SeDuMi. Even here, the angle grid would only have a one degree resolution which is not fine enough. Methods for angle grid refinement are on the other hand covered in [7] and other papers as well. Some implentational improvements could possibly alleviate the problem so that it would become feasible e.g. using parallel processing on the GPU, for example. Due to the lack of an implementation like this, ℓ_1-SVD is not used as a wideband method in the thesis.

4.4.2 Greedy methods

The greedy algorithms given for the narrowband problem in Section 3.3 can be extended to the wideband problem simply by evaluating the criterion of the best column such that it also takes the different subbands into account. The simplest way is similar to the incoherent methods (see Section 4.1) in the sense that the subbands are combined additively which corresponds to using the 1-norm. It would be possible to use also another norm like the euclidian norm. This was not tested, though, since the limit for the performance seems to be mostly restricted to the narrowband criterion discussed in Section 3.3. The additive combining (a method

similar to OMP-W) is used for the wideband problem in [8] and the algorithm we call OMP-W is originally presented in [23].

OOMP, which was described in Section 3.4.2, is extended to the wideband problem in Algorithm 9.

Input : The data $\boldsymbol{Y}_k$, angle grid $\vartheta_1, \ldots, \vartheta_M$, the set of used subband indices $\mathcal{K}$.

Initialize $\Omega^0 = \emptyset$, $\boldsymbol{r}_1(p,k)$ as the p'th column of $\boldsymbol{Y}_k$.

for $q = 1, \ldots, N$ *or until a stopping criterion is met* **do**

Denote $\Omega_m^q = \Omega^{q-1} \cup m$

$$m_{\min} = \underset{m \notin \Omega^{q-1}}{\arg\min} \quad \sum_{k \in \mathcal{K}} \left(\sum_{p=1}^{P} \| \boldsymbol{r}_q(p,k) - \boldsymbol{A}_{\Omega_m^q} \boldsymbol{A}_{\Omega_m^q}^{\dagger} \boldsymbol{r}_q(p,k) \|^2 \right)$$

$\Omega^q = \Omega^{q-1} \cup m_{\min}$

Update residuals

$$\boldsymbol{r}_{q+1}(p,k) = \boldsymbol{y}(p,k) - \boldsymbol{A}_{\Omega^q} \boldsymbol{A}_{\Omega^q}^{\dagger} \boldsymbol{y}(p,k)$$

end

Return Ω^q

Algorithm 9: OOMP-W

An efficient implementation of OMP is based on the use of QR decomposition $\boldsymbol{B}_k = \boldsymbol{Q}_k \boldsymbol{R}_k$. The orthogonal columns $\boldsymbol{\gamma}$ of $\boldsymbol{Q}$ are used in the algorithm summarized by Algorithm 10.

4.5 Other wideband methods

A wideband method that does not perform the subband decomposition but works on the sampled version of the model (5) directly was proposed in [24]. The method, however, requires inverse problems consisting of very large matrices which is why it is not used in the thesis. In the paper [24] a spline approximation is used for the source signals to overcome this problem but this is not possible for our case since no such assumption can be made from signals emitted from ships or submarines.

Input : The data $\boldsymbol{y}(p, \omega_k)$, angle grid $\vartheta_1, \ldots, \vartheta_M$, the set of used subband indices $\mathcal{K}$.

Initialize $\Omega^0 = \emptyset$, $\boldsymbol{r}_1(p,k) = \boldsymbol{y}(p, \omega_k)$.

for $q = 1, \ldots, N$ *or a stopping criterion is met* **do**

Choose best column

$$m_{\max} = \underset{m \notin \Omega^{q-1}}{\arg\min} \quad \sum_{k \in \mathcal{K}} \left(\sum_{p=1}^{P} \frac{|\langle \boldsymbol{a}(\vartheta_m, k), \boldsymbol{r}_q(p) \rangle|}{\|\boldsymbol{a}(\vartheta_m, k)\|} \right)$$

Update the set of selected indices

$$\Omega^q = \Omega^{q-1} \cup m_{\max}$$

Orthogonalize the basis vectors for all $k \in \mathcal{K}$

$$\boldsymbol{\gamma}_q(k) = \boldsymbol{a}(\vartheta_{m_{\max}}, k) - \sum_{l=1}^{q-1} \frac{\langle \boldsymbol{a}(\vartheta_{m_{\max}}, k), \boldsymbol{\gamma}_l(k) \rangle}{\|\boldsymbol{a}(\vartheta_{m_{\max}}, k)\|^2} \boldsymbol{\gamma}_l(k)$$

Update coefficients for all p and k

$$\beta_q(p,k) = \frac{\langle \boldsymbol{r}_q(p,k), \boldsymbol{\gamma}_q(k) \rangle}{\|\boldsymbol{\gamma}_q(k)\|^2}$$

Update residual for all p and k

$$\boldsymbol{r}_{q+1}(p,k) = \boldsymbol{r}_q(p,k) - \beta_q(p,k) \boldsymbol{\gamma}_q(k)$$

end

Set $\mathcal{S} = 0$ for all indices $m = 1, \ldots, M$.

Build the matrix $\boldsymbol{B}_k$ using the steering vectors corresponding to the indices Ω^q and set $\mathcal{S}(\Omega^q) = \sum_{p=1}^{P} \sum_{k \in \mathcal{K}} (\boldsymbol{B}_k^H \boldsymbol{B}_k)^{-1} \boldsymbol{B}_k^H \boldsymbol{y}(p, \omega_k)$

Output: $\mathcal{S}$

Algorithm 10: OMP-W

5 Likelihood based combining of subband estimates

The MVDR-S method presented in Section 4.1.1 is a simple way of combining the subbands to arrive at a wideband spectrum. Through experiments it was discovered that it suffers from some drawbacks: If the low frequencies, that have a poorer resolution ability, contain a lot of the signal energy, the resolved peaks in the high frequency subbands can become masked by the single wide lobes appearing in the low frequencies. This situation is illustrated in Figure 12.

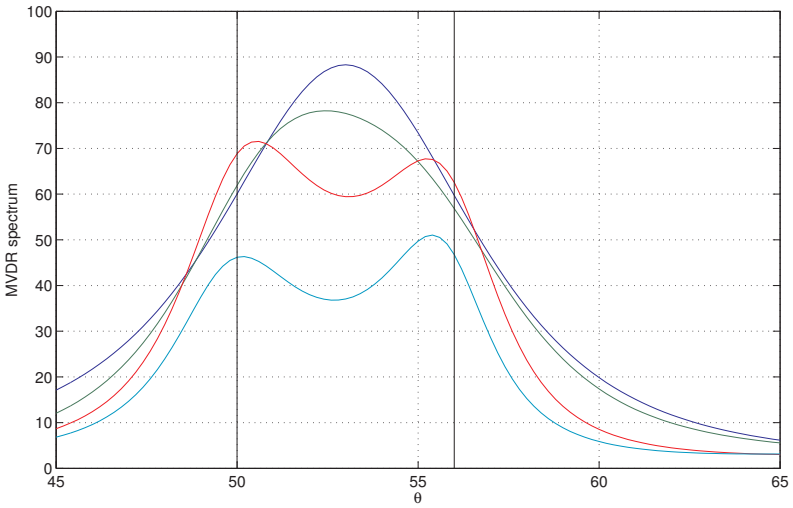


Figure 12: An example of a situation where the subband information is contradicting.

The MVDR estimator is also known to be biased. If, for example, there are two sources, the locations of the peaks become on average closer to each other than they should. This is also visible in Figure 12 – when two peaks are seen they are 'inside' the true angles. The different bias at different frequencies also affects the final combined result of MVDR-S.

The lobes of the MVDR spectra are typically also wider than those of MUSIC for example. Without a peak extraction procedure, this makes the combined spectrum look very flat and fuzzy which is not ideal when the spectrum is used visually for tracking the sources.

These drawbacks gave us the idea of replacing the conventional MVDR spectra with a likelihood based on the extracted peaks. This idea is now presented such that first we look at the statistical properties of MVDR and then describe the method itself. The idea is published in [3].

5.1 Statistical properties of MVDR

In order to build the likelihood function that can be used to combine the subbands, we need to know the probability distribution of a narrowband estimate. In other words, we need to know the type of the probability distribution and its parameters. Fortunately, the statistical properties of MVDR have been studied by Vaidyanathan and Buckley in [25] which is helpful. The publication [25] states that the estimated covariance matrix used in the MVDR estimation has a Wishart distribution with P degrees of freedom and that the inverse of this matrix has the inverse Wishart distribution. It also develops approximate formulas for the bias and variance of the estimated angles. We use these formulas and the normal distribution to build the likelihood function. Next, these approximate formulas and the reasoning behind them is given.

The denominator of the MVDR spectrum estimator

$$f(\theta, \boldsymbol{R}^{-1}) = \boldsymbol{a}^H(\theta)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta) \tag{84}$$

is chosen to be the function of interest. Naturally, at the peak location, the derivative (with respect to θ) of this function is zero:

$$\dot{f}(\tilde{\theta}_q, \boldsymbol{R}^{-1}) = 0. \tag{85}$$

The approximate formula for the asymptotic bias $\Delta\theta_q = \tilde{\theta}_q - \theta_q$ is derived using the first order Taylor series expansion around θ_q:

$$\dot{f}(\tilde{\theta}_q, \boldsymbol{R}^{-1}) \approx \dot{f}(\theta_q, \boldsymbol{R}^{-1}) + \ddot{f}(\theta_q, \boldsymbol{R}^{-1})\Delta\theta_q \tag{86}$$

which combined with (85) yields

$$\Delta\theta_q \approx -\frac{\dot{f}(\theta_q, \boldsymbol{R}^{-1})}{\ddot{f}(\theta_q, \boldsymbol{R}^{-1})} \tag{87}$$

where the expressions for the first and second order derivatives are

$$\dot{f}(\theta_q, \boldsymbol{R}^{-1}) = 2\,\mathrm{Re}[\dot{\boldsymbol{a}}^H\boldsymbol{R}^{-1}\boldsymbol{a}] \tag{88}$$

$$\ddot{f}(\theta_q, \boldsymbol{R}^{-1}) = 2\,\mathrm{Re}[\dot{\boldsymbol{a}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{a}} + \boldsymbol{a}^H\boldsymbol{R}^{-1}\ddot{\boldsymbol{a}}]. \tag{89}$$

Denoting $C = -j2\pi f x_n/c$ the derivatives of the steering vectors are

$$\dot{a}_n(\theta) = -C\sin\theta e^{C\cos\theta} \tag{90}$$

and

$$\ddot{a}_n(\theta) = (-C\cos\theta + (C\sin\theta)^2)e^{C\cos\theta}. \tag{91}$$

The derivation of the variance is done similarly just expanding around the biased estimate. The variance $\sigma^2_{q,k} = E[(\hat{\theta}_{q,k} - \Delta\theta_{q,k} - \theta_{q,k})^2]$ can be approximated with

$$\sigma^2_{q,k} \approx \frac{(P-N)}{(P-N-1)(P-N+1)} \frac{\mathrm{Re}[\boldsymbol{a}^H \boldsymbol{R}_k^{-1} \boldsymbol{B} \boldsymbol{R}_k^{-1} \dot{\boldsymbol{a}}]}{\ddot{f}^2}, \tag{92}$$

where $\boldsymbol{B} = \dot{\boldsymbol{a}}\boldsymbol{a}^H + \boldsymbol{a}\dot{\boldsymbol{a}}^H$.

This far, it has been assumed that the true covariance matrix is known. Let us recall its structure: Let $\boldsymbol{A}$ be the matrix having as columns the steering vectors of all the sources. Assuming the sensors signals contaminated by white noise with variance $\sigma^2_{e,k}$, their covariance matrix has the expression

$$\boldsymbol{R} = \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^H + \sigma_e^2 \boldsymbol{I}, \tag{93}$$

where $\boldsymbol{S}$ is the covariance matrix of source signals.

The proposition is to replace this true covariance matrix with the estimation

$$\boldsymbol{R}_k \approx \sum_{q=1}^{q_k} P_k(\hat{\theta}_{q,k}) \boldsymbol{a}(\hat{\theta}_{q,k}, f_k) \boldsymbol{a}^H(\hat{\theta}_{q,k}, f_k) + \hat{\sigma}^2_{e,k} \boldsymbol{I}, \tag{94}$$

which means that in (93) we assume that $\boldsymbol{S}_k$ is diagonal (uncorrelated sources). The power of the sources and the noise power are given by Algorithm 11.

With this estimation the bias (87) and variance (92) can be evaluated. Since, in addition, we have made the assumption that the angles are normally distributed (which is confirmed by our simulations), the whole probability distribution for a DOA estimate is known. It can be said that the random variable θ_q follows the normal distribution that is corrected for the bias (87) and has the variance (92) i.e

$$\theta_q \sim \mathcal{N}(\theta_q - \Delta\theta_q, \sigma_q^2). \tag{95}$$

Figure 13 illustrates the probability distribution around the peak of the MVDR spectrum.

Input : snapshots $Y(p, \omega_k)$, $p = 1, \ldots, P$, $k \in \mathcal{K}$
angle grid ϑ_m, $m = 1, \ldots, M$
SNR threshold $\mathrm{SNR}_{\mathrm{TH}}$

for $k \in \mathcal{K}$ **do**
  Estimate $\hat{\boldsymbol{R}}$ like in (25)
  for $m = 1, \ldots, M$ **do**
    $P_k(\vartheta_m) = \left(\boldsymbol{a}^H(\vartheta_m, \omega_k) \hat{\boldsymbol{R}}^{-1} \boldsymbol{a}(\vartheta_m, \omega_k) \right)^{-1}$
  end
  Pick the local maxima of $P_k(\vartheta_m)$ and denote $\hat{\theta}_{1,k}, \ldots, \hat{\theta}_{q_k,k}$ the locations of the maxima and $P_k(\hat{\theta}_{1,k}), \ldots, P_k(\hat{\theta}_{q_k,k})$ the values of the maxima
  Estimate the power of noise $\hat{\sigma}_{e,k}$ as the average of the $N - \hat{Q}$ smallest eigenvalues of $\hat{\boldsymbol{R}}$ using the MDL criterion (44) for $\hat{Q}$
  for $q = 1, 2, \ldots, q_k$ **do**
    if $10 \log_{10} \frac{P_k(\hat{\theta}_{q,k})}{\hat{\sigma}^2_{e,k}} < SNR_{TH}$ **then**
      Discard the q-th peak because the SNR is below threshold.
    end
  end
end

Output: The DOA estimates $\hat{\theta}_{1,k}, \ldots, \hat{\theta}_{q_k,k}$, the powers $P_k(\hat{\theta}_{1,k}), \ldots, P_k(\hat{\theta}_{q_k,k})$ and $\hat{\sigma}_{e,k}$ for all $k \in \mathcal{K}$.

Algorithm 11: Subband estimations with MVDR.

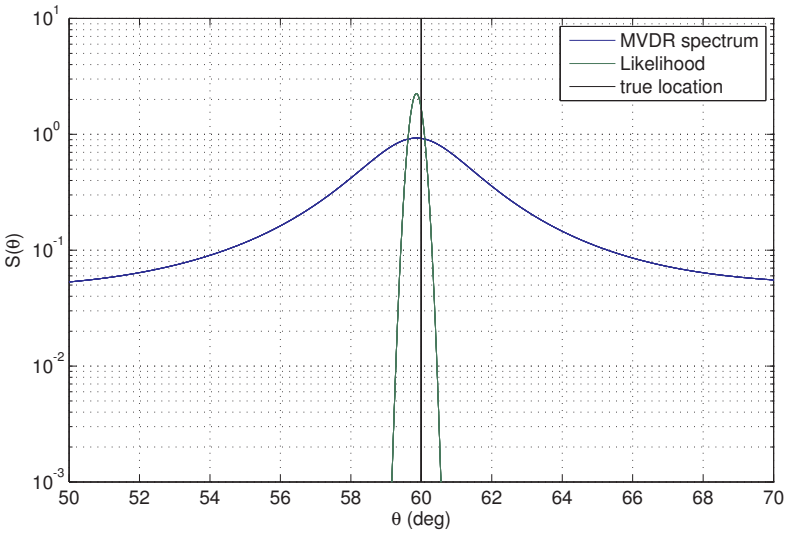


Figure 13: An example of the MVDR spectrum and the probability distribution function relating to the location of the peak. The source is at 60°.

5.2 Proposed method

Based on the previous discussion, we propose to estimate the wideband DOAs using a likelihood function combining for all subbands the probabilities that a DOA falls on a certain small interval. The likelihood function is built using the MVDR spectrum information given by Algorithm 11 and the bias and variance estimations (87) and (92) for each subband. More specifically, given a finite grid of angles ϑ_m, $m = 1 : M$, we estimate the log-likelihood

$$L(\vartheta_m) = \log \Pr(\theta \in [\vartheta_{m-1}, \vartheta_m]) \tag{96}$$

with

$$L(\vartheta_m) = \log \prod_{k \in \mathcal{K}} \Pr(\theta \in [\vartheta_{m-1}, \vartheta_m] \mid \hat{\theta}_{q,k}). \tag{97}$$

By $\hat{\theta}_{q,k}$ we denote the peak of the MVDR spectrum in subband k that is closest to ϑ_m. The probabilities from (97) are evaluated as

$$\Pr(\theta \in [\vartheta_{m-1}, \vartheta_m] \mid \hat{\theta}_{q,k}) = \Gamma(\vartheta_m) - \Gamma(\vartheta_{m-1}) \tag{98}$$

where Γ is the cumulative distribution function (cdf) of a mixture of the gaussian and the uniform distribution:

$$\Gamma(\theta) = \beta \Phi_{\hat{\theta}_{q,k} - \Delta\theta_{q,k}, \sigma_{q,k}}(\theta) + (1 - \beta) U_{[0,180^\circ]}(\theta). \tag{99}$$

The normal cdf $\Phi_{\hat{\theta}_{q,k} - \Delta\theta_{q,k}, \sigma_{q,k}}(\theta)$ is corrected for the bias (87) and has the variance (92).

The uniform cdf U in (99) is effectively used as a hard threshold to prevent extremely small probabilities that would practically annihilate (97). In this manner, we can simply take into account subbands where the SNR is too low or where a source is not active. The overlapping area of the uniform distribution and the normal distribution is typically small which is why we have chosen $\beta = \frac{1}{2}$.

Algorithm 12 summarizes the operations described above for combining the narrowband estimations into a single wideband likelihood function. We name the algorithm MVDR-LBC.

A possible refinement of the algorithm is to take into account that the subbands signals emitted by a source are not sinusoidal, as they are assumed in (93) and (94), since they are obtained by FFT filtering of wideband signals. A simple approximation is to consider that the sources have a constant spectrum on the interval $[f_k - \Delta f, f_k + \Delta_f]$, where $\Delta_f = f_s/4K$, and are not active elsewhere. (This amounts

Input : estimated DOAs $\hat{\theta}_{1,k}, \ldots, \hat{\theta}_{q_k,k}$
powers $P_k(\hat{\theta}_{1,k}), \ldots, P_k(\hat{\theta}_{q_k,k})$
noise variances $\hat{\sigma}_{e,k}$, $k = 1, \ldots, K$
angle grid ϑ_m, $m = 1, \ldots, M$
significant bands $\mathcal{K} \subset \{1, \ldots, K\}$

for $m = 1, \ldots, M$ **do**
  Initialize log-likelihood $L(\vartheta_m) = 0$
  for $k \in \mathcal{K}$ **do**
    1. Build $\boldsymbol{R}_k$ as in (94)
    2. Take the closest DOA $\hat{\theta}_{q,k}$ to ϑ_m and use the corresponding steering vector and eqs. (87), (92) to estimate bias $\Delta\theta_{q,k}$ and variance $\sigma_{q,k}$
    3. Using (98) and (99) evaluate $\varpi_k = \Pr(\theta \in [\vartheta_{m-1}, \vartheta_m] \mid \hat{\theta}_{q,k})$
    4. Update $L(\vartheta_m) = L(\vartheta_m) + \log \varpi_k$
  end
end

Output: $L(\vartheta_m)$, $m = 1, \ldots, M$

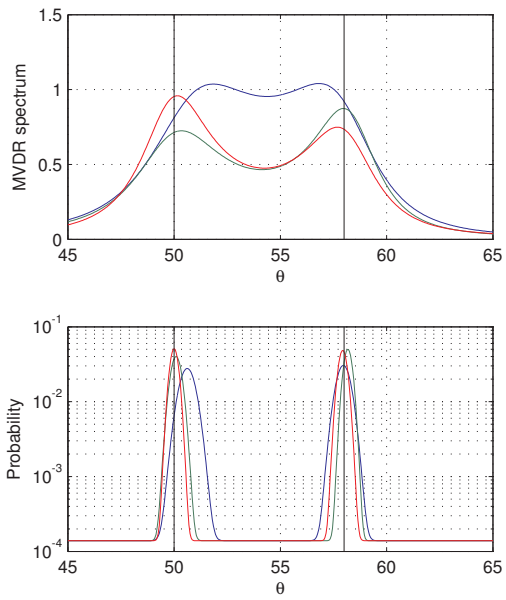
Algorithm 12: Likelihood Based Combining of Subband Estimates using MVDR as the subband estimator. The algorithm is called MVDR-LBC.

to replacing FFT with an ideal bandpass filter.) In this case, (94) is replaced with

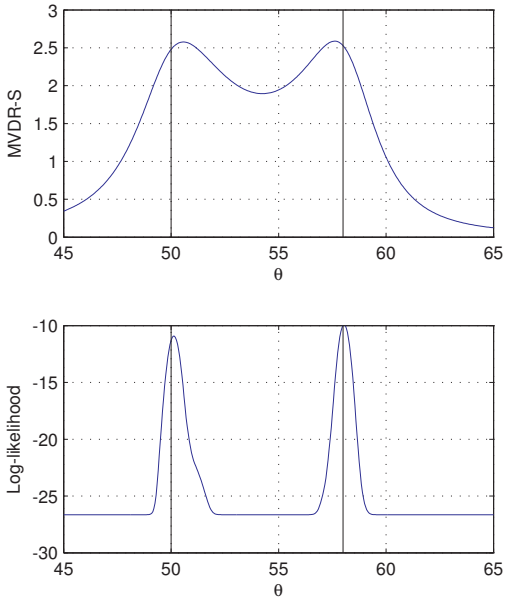
$$\boldsymbol{R}_k \approx \sum_{q=1}^{q_k} P_k(\hat{\theta}_{q,k}) \int_{f_k-\Delta f}^{f_k+\Delta f} \boldsymbol{a}(\hat{\theta}_{q,k}, f)\boldsymbol{a}^H(\hat{\theta}_{q,k}, f)df + \hat{\sigma}_{e,k}^2 \boldsymbol{I}, \tag{100}$$

The integral above can be easily solved analytically. At least in the simulation setup used in Section 8, the difference between (94) and (100) is small and hence we have systematically used (94) due to its faster computation.

Figure 14 illustrates the difference between MVDR-S and MVDR-LBC. The upper left figure contains the MVDR spectrum estimation for a few frequency bands where two sources can be distinguished. The lower left figure shows the probability density functions corresponding to (99). The rightmost figures show the combined results using MVDR-S and MVDR-LBC. It can be seen that by compensating the bias due to close sources it is possible to achieve a more accurate estimation. Also, the probability distribution functions are much sharper, promising better possibilities of close sources separation. When the information over the subbands is contradicting like in Figure 12 the sharpness at least lets the true source locations to still be visible in the wideband spectrum and not become masked by the low frequencies.



(a) The subband spectra for 1687.5 Hz (blue), 1937.5 Hz (green) and 2000 Hz (red). Upper: the MVDR spectra. Lower: The likelihoods of the extracted peaks.



(b) The combined spectra. Upper: the summation based MVDR. Lower: the likelihood based combining.

Figure 14: An example of the difference between MVDR-S and MVDR-LBC. The black vertical lines show the locations of the true sources.

6 Sparse autoregressive modeling of underwater signals

In general, a model is used to explain the physical system that generates a signal. In this section, the data recorded with a hydrophone in underwater is modeled with a specific type of stochastic model – namely with a sparse autoregressive model. First the basic model is explained, then a practical method for using it, the linear prediction, is intoduced and then extended to the sparse case via a specific algorithm. Finally, a simple lossless compression method that can be used to measure the goodness of the model is described.

In a stochastic model, the observations are explained as the output of a discrete time linear filter that is fed by a purely random process [26, p. 45]. In general, the filter in the stochastic model, which is illustrated by Figure 15, can be of whatever type: a finite impulse response (FIR) filter or an infinite impulse response (IIR) filter. In the case of the *autoregressive model* the filter is a specific type IIR filter: the output of it depends on the past values of the output and the current value of the input but not on the previous input values. As a difference equation this means

$$u(t) = w_1 u(t-1) + w_2 u(t-2) + \cdots + w_L u(t-L) + v(t) \tag{101}$$

where u denotes the autoregressive process and v the random input process. Denoting the z-transforms of them by capital letters, we can write

$$U(z) = \frac{1}{1 - \sum_{l=1}^{L} w_l z^{-l}} V(z). \tag{102}$$

The coefficients w_l are known as the AR parameters and L is known as the model order. The complementary method for using the autoregressive model to determine the AR parameters is known as linear prediction. Before moving into prediction let us review the well known matrix equation for estimating the AR parameters called the Yule-Walker equations. They are built as follows: First the equation (101) is multiplied with $u(t-k)$ and then the expection operator is applied to both sides:

$$\mathrm{E}[u(t)u(t-k)] = \sum_{l=1}^{L} w_l \,\mathrm{E}[u(t-l)u(t-k)] + \mathrm{E}[v(t)u(t-k)]. \tag{103}$$

The rightmost term is nonzero only for $k = 0$ since only the current output value is dependent on the random input signal. The expectation of a signal multiplied with a delayed value of itself is known as the *autocorrelation function* and is denoted by

r. Thus, the previous equation can be written for $k > 0$ as

$$r(k) = \sum_{l=1}^{L} w_l r(l-k) \tag{104}$$

when this process is repeated for $k = 1, \ldots, L$ and noting that $r(k) = r(-k)$ we arrive at the matrix equation called the Yule-Walker equations:

$$\begin{bmatrix} r(0) & r(1) & & r(L-1) \\ r(1) & r(0) & & r(L-2) \\ & & \ddots & \\ r(L-1) & r(L-2) & & r(0) \end{bmatrix} \begin{bmatrix} w(1) \\ w(2) \\ \vdots \\ w(L) \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(L) \end{bmatrix} \tag{105}$$

or

$$\boldsymbol{\Phi w} = \boldsymbol{z} \tag{106}$$

in short.

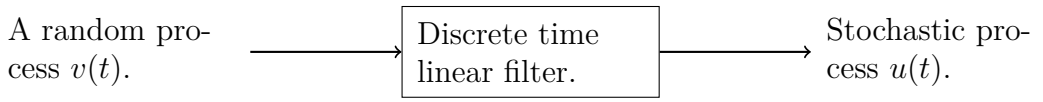


Figure 15: Stochastic model.

A simple autoregressive model where L is quite small (e.g. 10) is considered to be a good model for the vocal tract and thus can be used for speech signals. This has been done already in 1966 by Saiko and Itakura [26, p. 30]. For lossless audio compression this is also a useful strategy and used by FLAC [27], for example. However, a simple model like this is only good for predicting the spectral envelope of the signal. The harmonic content can not be taken into account because it would require a long delay term in (101). A separate long term predictor can be used for speech by estimating the pitch.

For music signals, however, a better approach seems to be to have a sparse model (predictor) which means that in (101) L can be very long (e.g. 500) but most of the parameters w_l are equal to zero.

Giacobello et al. talk about sparse audio predictors for audio in [28]. They illustrate the idea that a harmonic audio signal consists of the pitch and the formant. The latter can be modeled with the non-sparse predictor while the pitch can only be modeled with a long term predictor. The combination of these can be viewed as the sparse predictor. A more extensive study about modeling audio and music with sparse predictors can be found in [29]. In it, it was discovered that when the non-sparse predictor leaves periodicities in the residual signal, the sparse predictor outperforms the non-sparse.

6.1 Adaptive prediction

Predicting refers to the process of using the previous samples of a signal to predict the current sample. Denoting the *prediction order* by L, the estimate for the sample at time instant t is

$$\hat{u}(t) = w_1 u(t-1) + w_2 u(t-2) + \cdots + w_L u(t-L) = \sum_{l=1}^{L} w_l u(t-l) \tag{107}$$

which is depicted in Figure 16. The prediction error, also known as the *residual* is the difference between the true sample and the prediction:

$$e(t) = u(t) - \hat{u}(t). \tag{108}$$

The predictor weights $w_1, \ldots, w_L$ can be designed via least squares applied to the equation

$$\boldsymbol{u} = \boldsymbol{M}\boldsymbol{w} \tag{109}$$

$\boldsymbol{M}$ being the $N \times L$ Toeplitz *data matrix*

$$\boldsymbol{M} = \begin{bmatrix} u(t-1) & 0 & & 0 \\ u(t-2) & u(t-1) & & 0 \\ & & \ddots & \\ u(t-L) & u(t-L+1) & & u(t-1) \\ \vdots & \vdots & & \vdots \\ u(t-N) & u(t-N+1) & & u(t-N+L-1) \end{bmatrix}. \tag{110}$$

N refers to the length of the frame i.e. the number of data samples used in designing the predictor weights where $N >> L$. The least squares solution yields

$$\hat{\boldsymbol{w}} = (\boldsymbol{M}^H \boldsymbol{M})^{-1} \boldsymbol{M}^H \boldsymbol{u}. \tag{111}$$

It can be noted that $\boldsymbol{M}^H\boldsymbol{M}$ is in fact the sample estimate for the autocorrelation matrix $\boldsymbol{\Phi}$ and $\boldsymbol{M}^H\boldsymbol{y}$ is the estimate for $\boldsymbol{z}$ so, in fact, the least squares solution is the solution to the Yule-Walker equations (106) using the sample estimates for the autocorrelation. This method can be used directly when there is a fixed frame of data (input audio) that is processed as a batch and is the conventional method used in FLAC, for example. We, however, use an adaptive filtering scheme where the predictor $\boldsymbol{w}$ is updated at each time instant.

Recursive Least Squares (RLS) is a standard adaptive filtering algorithm. The adaptive nature means that the filter is updated at each time instant according to the error of the output of the filter and the desired output. In the prediction context the

desired signal is the current signal sample $u(t)$ and the input signal for the adaptive algorithm is the one-step delayed version of the signal $\boldsymbol{u}(t-1)$. The situation is depicted in Figure 17.

The idea behind RLS is the same as in the batch least squares estimation. At each time instant the prediction error is to be minimized in the least squares sense. However, in RLS, an additional, regularization term is introduced into the criterion in order to smooth the solution $\boldsymbol{w}$. The criterion to be minimized in RLS is

$$\mathcal{J}(t) = \sum_{i=1}^{t} \lambda^{t-i} |u(t) - \boldsymbol{w}(t)^H \boldsymbol{u}(t-1)|^2 + \delta \lambda^t \|\boldsymbol{w}(t)\|^2 \tag{112}$$

where λ is the forgetting factor used to forget data in the distant past and δ is a regularization parameter. Now let us compare this to the normal equations of the least squares solution which was found to similar to the Yule-Walker equations. The additional term changes the autocorrelation matrix such that it becomes 'diagonally loaded' i.e.

$$\Phi(t) = \sum_{i=1}^{t} \lambda^{t-i} \boldsymbol{u}(t-1)\boldsymbol{u}(t-1)^H + \delta \lambda^t \boldsymbol{I} \tag{113}$$

and the autocorrelation vector on the right hand side becomes

$$z(t) = \sum_{i=1}^{t} \lambda^{t-i} \boldsymbol{u}(t-1)u(t) \tag{114}$$

This equation

$$\boldsymbol{\Phi}(t)\boldsymbol{w}(t) = \boldsymbol{z}(t) \tag{115}$$

is what RLS solves at each time instant t. In order to process efficiently, RLS updates the autocorrelations as follows:

$$\boldsymbol{\Phi}(t) = \lambda \Phi(t-1) + \boldsymbol{u}(t-1)\boldsymbol{u}^H(t-1) \tag{116}$$

$$\boldsymbol{z}(t) = \lambda z(t-1) + \boldsymbol{u}(t-1)u(t) \tag{117}$$

The basic implementation of RLS uses the so called matrix inversion lemma to find a simple expression for $\boldsymbol{\Phi}^{-1}(t)$ that is needed to solve $\boldsymbol{w}(t)$ and by doing so proceeds to iterate over the time samples.

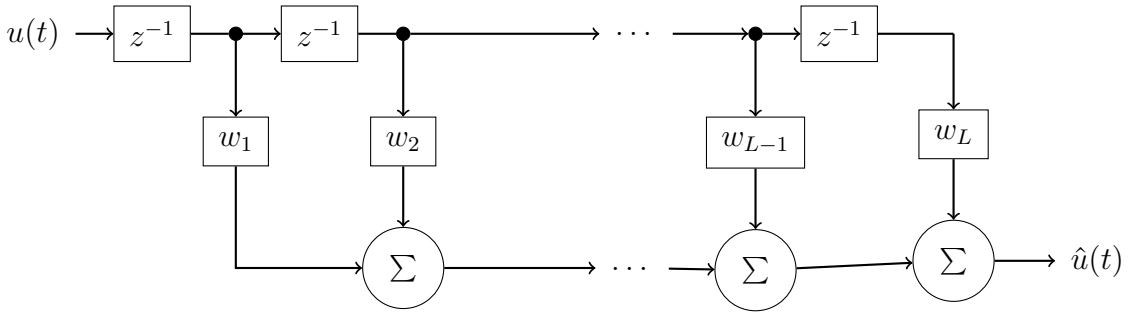


Figure 16: Prediction.

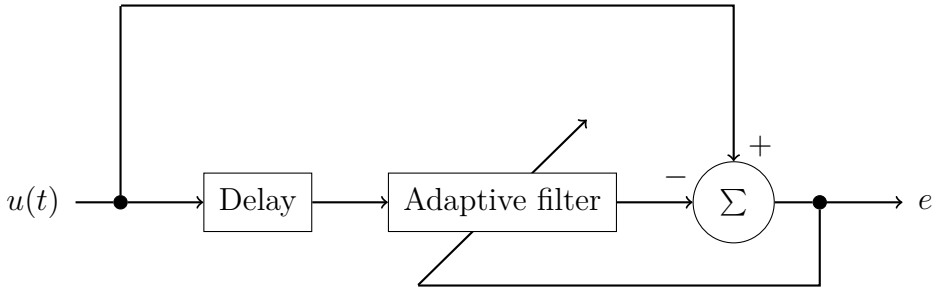


Figure 17: Prediction using an adaptive filter.

6.2 Introducing sparsity via GRLS

In the previous section a method for updating the values of the predictor $\boldsymbol{w}$, RLS, was introduced. There are, however, two more problems that need solving: Determining the order and the locations of nonzero coefficients. For the regular, non-sparse predictor only the optimal number of coefficients L, which is known as the order, needs to be determined. This can be done using an information theoretic criterion such as AIC or MDL.

When the predictor is sparse, i.e. most of its coefficients are equal to zero, the number of nonzero elements need to be determined as well as the locations of the nonzeros. Similarly to Section 3.3 this could be formulated as an ℓ_0-optimization problem which can be solved efficiently with the basis pursuit approach that solves the ℓ_1 problem instead. Under the research done concerning sparse audio predictors it was noted that the more tractable greedy approaches (OMP and OOMP) work well for the problem. The greedy algorithm can thus be used to determine the locations of the nonzeros such that they are naturally ordered with respect to the goodness of the location. The number of nonzeros i.e. when to stop iterating the greedy algorithm can be determined again with an information theoretic criterion.

Greedy Sparse RLS (GRLS) is a method introduced in [11] that performs at each time instant both the coefficient estimation step done in basic RLS and a basis selection step. Basis refers to the nonzero positions of the sparse predictor. For the

details of GRLS, the reader is directed to [11] but the basic principle of it is the following: GRLS is based on the OLS (the essence of which is known in this thesis as OOMP, see Section 3.4.2) algorithm and thus maintains a QR decomposition of the data matrix $\boldsymbol{M}$. By using orthogonal operators such as Givens Rotations at each time instant t if performs the following:

1. Solve the predictor weights $\boldsymbol{w}$ using the same criterion as in the standard RLS and updates the residual.

2. Choose the basis by allowing neighboring columns of the data matrix (which correspond to different delays) to change at every time instant. For the last selected column, it allows all of the non-active columns to compete. All of this is accomplished using the same criterion as in OLS.

The GRLS article [11] also gives a number of information theoretic criteria to select the number of nonzero positions which in the sparse context is equivalent to the model order. In our experiments we use the so called Predictive Least Squares (PLS) criterion, see [11] for details.

It is of interest to see how the sparse predictors work compared to a non-sparse one. The non-sparse prediction using RLS is used in MPEG-4-ALS [30], for example. The same algorithm – GRLS – is used for both, the sparse and the non-sparse one. For the non-sparse predictor the basis selection step is however omitted. The order of the non-sparse predictor is chosen with PLS. This operation is essentially the same as in standard RLS.

In Section 8 is is shown that for some data the sparse predictor can improve the compression and thus is a better model than the non-sparse one. To illustrate the difference between the sparse and the non-sparse predictor a set of examplar figures are given resulting from the processing of a hydrophone recording. The predictors themselves are shown in Figure 18. It can be noticed that the largest weights are in both cases in the beginning. While the non-sparse predictor have quite large weights also in the range 10–50 the sparse predictor is already equal to zero in many of these delay values. Instead, the sparse predictor finds large weights in the range around the delay index 400.

The effect of the prediction can be visualized in the frequency domain by using the parametric spectrum given by placing $z = e^{j\omega}$ into (102) and using the esimates for $\boldsymbol{w}$. The amplitude response i.e. $|U(e^{j\omega})|$ is shown in Figure 19 and it is compared to the periodogram of the signal computed by using the previous 2000 samples of the signal. It is very obvious that the long-term prediction provided by the sparse predictor can mimic the harmonic shape present in the original signal while the non-sparse one only gives the spectral envelope.

Like [29] states, the sparse predictor becomes beneficial when the autocorrelation of the non-sparse residual signal still contains clear peaks which imply that there are still harmonic content present in the residual signal. Continuing our example, the estimated autocorrelations using the last 2000 samples of the original signal, the non-sparse residual and the sparse residual are shown in Figure 20. The non-zero positions of the sparse predictor, which can also be seen from Figure 18, are also shown there. The most obvious peak of the non-sparse residual signal is at the lag value 400. This is also the point where the sparse predictor finds non-zero values which is to show that the method, GRLS, works.

The order selected by the same PLS criterion for both the sparse and non-sparse prediction is visualized for the example in Figure 21. The order for the sparse predictor is smaller than for the non-sparse predictor. This is quite consistent and can be seen also in Section 8. The need for fewer parameters (non-zero weights) with the sparse model can reduce complexity and promises intuitively better modeling.

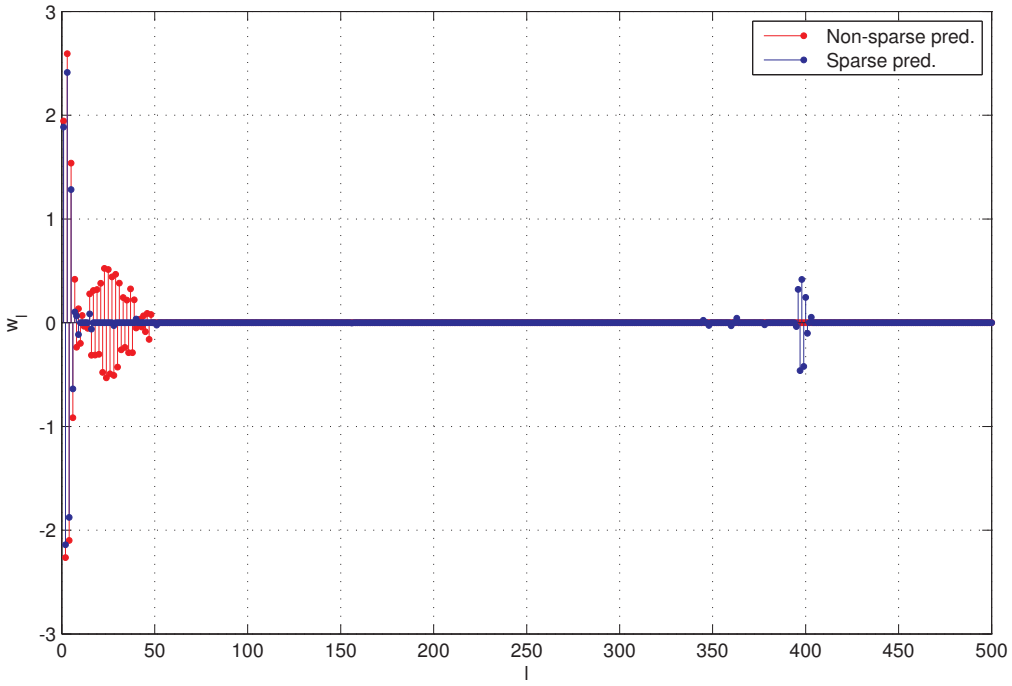


Figure 18: The predictors.

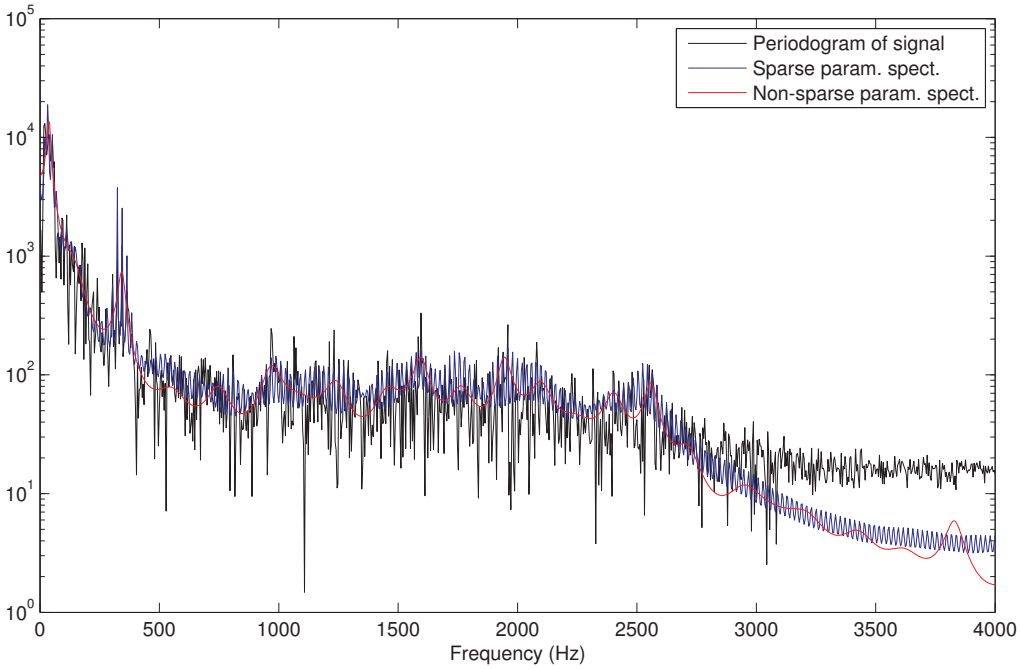


Figure 19: The spectra for a hydrophone file.

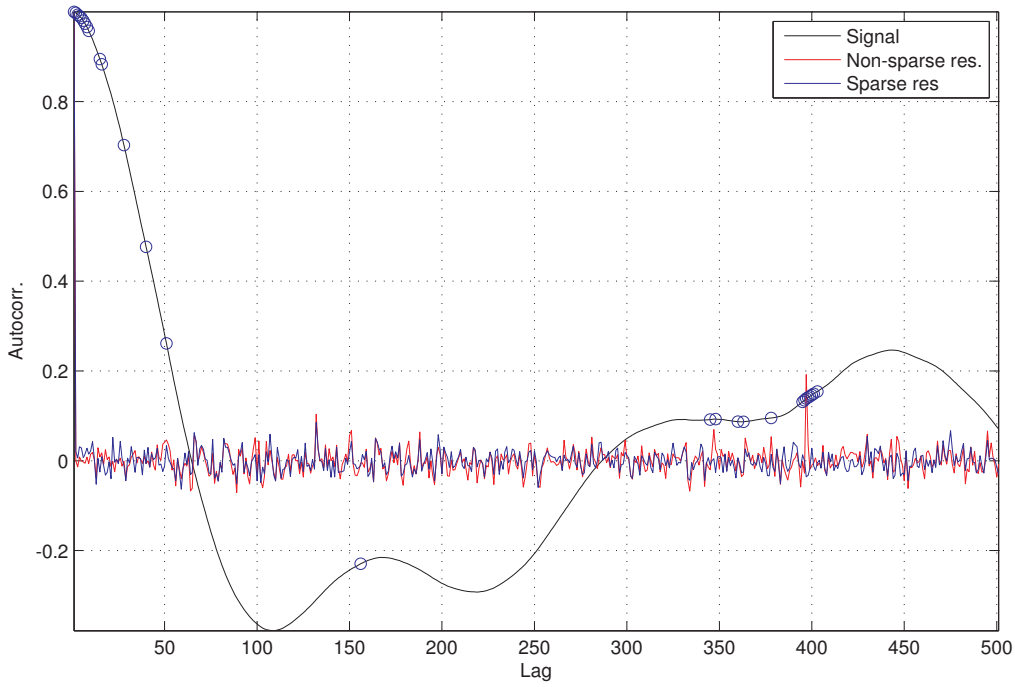


Figure 20: The estimated autocorrelation function for a hydrophone file. The blue circles denote the locations of the sparse predictor.

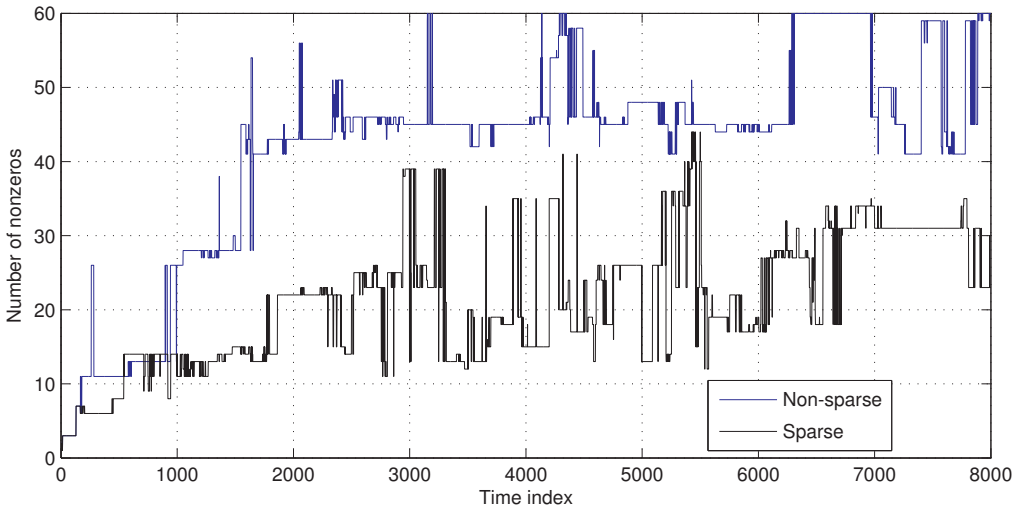


Figure 21: The order stays lower for the sparse method.

6.3 Lossless compression with Golomb-Rice entropy coding

The linear prediction acts as a method for using the autoregressive model. It whitens the input signal (recorded hydrophone audio) and by such reduces the redundancy between the time samples of the input signal. The values of the output signal (residual) are small compared to the input signal and can be coded with an entropy coder to produce a bitstream. The code length of the bitstream can be measured. A shorter code length means better compression. Since we have two alternative methods for the linear prediction, the sparse and the non-sparse one, they can be compared by using the same entropy coder and measuring the code length. If the code length is consistently shorter for one of the methods it is possible to conclude that the model behind it is better.

For the entropy coding we choose to use a method called the Golomb-Rice entropy coder. This simple, yet effective method is used in Shorten, FLAC and MPEG4-ALS among others and was first introduced by Golomb [31] and further developed by Rice [32]. The Golomb-Rice is in fact an optimal prefix code when the input follows the geometric distribution. Our implementation of the Golomb-Rice encoder is described in Algorithm 13. The total process of encoding the hydrophone signals, the samples of which are 16 bit integers, is thus the following

1. Use GRLS to produce the residual and round the value to an integer.
2. Use the Golomb-Rice encoder given in Algorithm 13 to write the bitstream.

This process can be reverted with a symmetric Golomb-Rice decoder and GRLS to decode the signals.

Input : A frame of residual samples $\boldsymbol{e}$.

1 Calculate the code length for $M = 0, 1, \ldots, 15$ and choose the smallest.

2 Write the optimum M into the bitstream as an unsigned binary code. This takes 4 bits.

3 **for** *each* t **do**

4 Compute the quotient

$$q = \left\lfloor \frac{|e(t)|}{2^M} \right\rfloor$$

and write it as an unary code with symbol '1' into the bitstream. This takes q bits.

5 Add a symbol '0' into the bitstream to mark boundary.

6 Take the remainder

$$r = |e(t)| - q \cdot 2^M$$

and use M bits to write it into the bitstream using binary code.

7 **if** $e(t) \neq 0$ **then**

8 Write '0' into the bitstream if $e(t) < 0$ and a '1' if $e(t) > 0$.

9 **end**

10 **end**

Algorithm 13: Golomb-Rice encoder.

7 Peculiarities of underwater surveillance

This section focuses on the sea environment and the peculiar aspects of underwater surveillance. The sensor array used in the results is described. The acoustic signatures of ships, propagation of sound in the sea and the concept of real time tracking is covered.

7.1 The towed sensor array

The system used in the sea to capture the data consists of an *unmanned underwater vehicle* (UUV) and the array of hydrophones towed by it. The setup is illustrated in Figure 22. A hydrophone is the equivalent of a microphone in underwater. They can be thought of as being omnidirectional i.e. the response of the hydrophone is a perfect sphere in three dimensions. A hydrophone is inherently an analogous device which is why electronics (an ADC for example) is needed to convert to a digital signal. We assume that each hydrophone has similar electronics which is why there is no need of having to pay extra attention to the electronics.

The hydrophones can be attached to an elastic, rope-like material. One end of this rope is attached to the UUV and the other can be lifted by a buoy for example. The whole system is in general moving. This forward movement, gravity and the possibly elastic material is prone to affect the exact locations of the hydrophones which can cause modeling errors. These kind of errors could be taken into account in the DOA estimation methods also. A robust version of MVDR, for example, was proposed in [33].

The hydrophone array used throughout this thesis is a uniform linear array with $N = 18$ sensors with a distance $d = 0.25$ m between them.

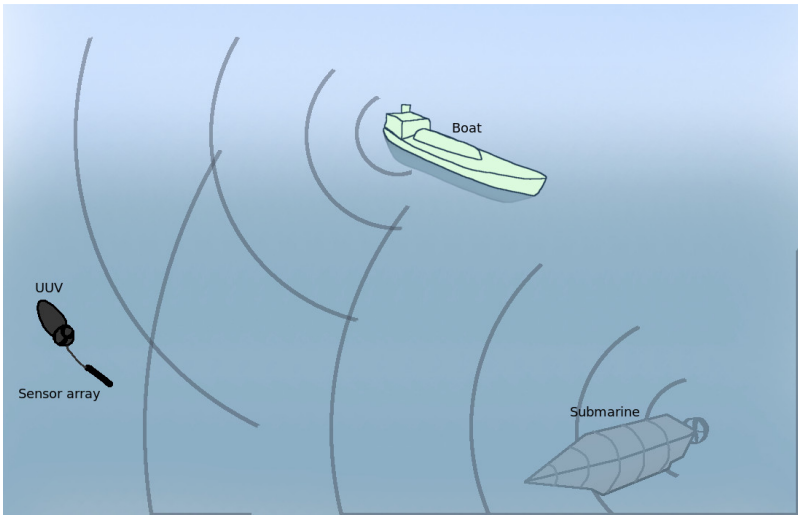


Figure 22: Basic setup

It is well known that the sensor array for narrowband signals should be designed so that the spacing between two sensors equals to half of the wavelength of the arriving signal i.e.

$$d = \frac{\lambda}{2} = \frac{c}{2f}. \tag{118}$$

This is the spatial equivalent of the Nyquist sampling rate and guarantees that the arriving signal is not aliasing. Decreasing the sensor spacing would weaken the ability to resolve close sources. For wideband signals this means that the sensor array is optimal in this sense only for a single frequency. Lower frequencies suffer from worse resolution while higher are prone to aliasing. This means that the highest frequency that can be received with our sensor array without aliasing is $f_{\max} = \frac{d}{2c} = \frac{1450\,\mathrm{m/s}}{2 \cdot 0.25\,\mathrm{m}} = 2900$ Hz.

As mentioned in Section 3, the DOA methods used – MVDR and MUSIC – are superresolution methods which means that they can resolve sources that Beamformer could not. The resolution limit for Beamformer, also known as the beamwidth is, however, easy to compute and gives a rough idea about the effect of frequency on the resolution ability. According to [12] the beamwidth

$$\Delta\theta \approx \frac{\lambda}{x_N} \tag{119}$$

where λ is the wavelength of the arriving signal and x_N is the length of the array (see Section 2), in our case $x_N = 4.5$ m. The beamwidth is depicted in Figure 23 as a function of frequency. The array performs best for frequencies over 2000 Hz. At 500 Hz the resolution is already very poor and at 250 Hz the resolution ability has vanished completely.

In the wideband DOA literature this problem of having wide differences in the resolution ability from frequency band to another has not been paid much attention to. Let us look at the papers which address the close source problem and their experimental setup: [20] uses a setup which results in the beamwidth variation shown in Figure 25. [4] uses virtually the same setup, only with less sensors. [6] only gives proportional values instead of exact frequencies but also relies on a much more narrow frequency span than in our case. The variation of beamwidth is thus much larger in our case and causes for example MVDR to be more prone to have contradicting information in the subbands as was described in Section 5.

In the experimental Section 8 we describe results mostly for two different scenarios: One which uses all the information available i.e. frequencies starting from very low and one which restricts to high frequencies. The first is illustrated with the ideal frequency response of the filter bank that produces the subbands in Figure 23 and the second in Figure 24 respectively.

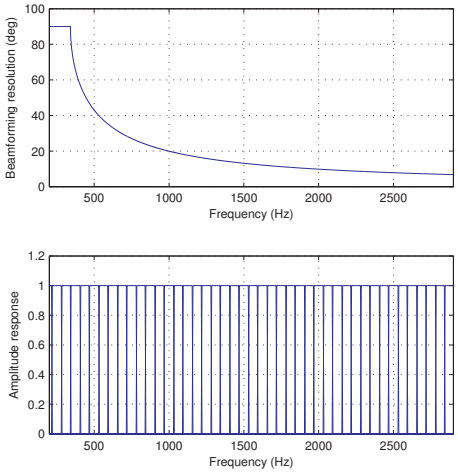


Figure 23: The beamwidth and the ideal amplitude response of the filter bank for all frequencies.

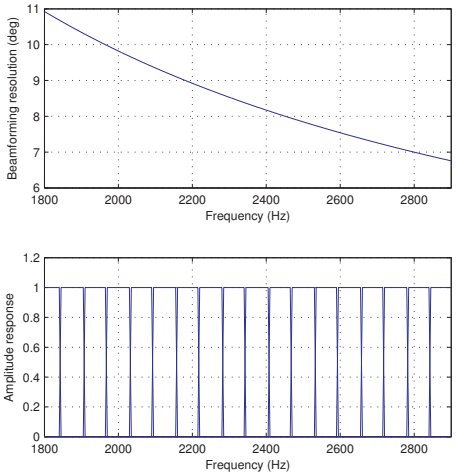


Figure 24: The beamwidth and the ideal amplitude response of the filter bank for high frequencies.

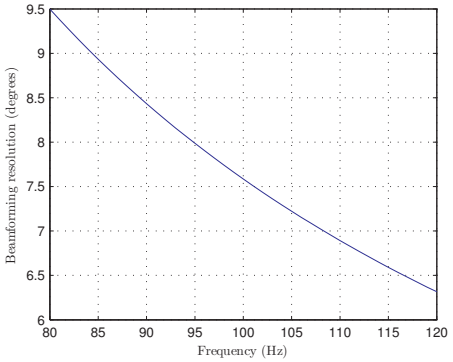


Figure 25: A comparative example of the beamwidth used in other publications.

7.2 The sea environment and the acoustic signature of ships

When sound propagates through a medium, such as water, it attenuates with growing distance from the source. This attenuation is known as absorption loss and in the sea it is mainly caused by the viscosity of water. It is clear, that the lower the frequency of sound the smaller the absorption loss is (at least until very low frequencies e.g. 100 Hz). In [34, p. 141] a plot is given which shows how the distance from the source and the frequency are related using a boundary of 6 dB absorbtion loss. It can be interpreted, for example, that a 2 kHz sound wave is attenuated 6 dB at 20 km distance from the source. Higher frequencies would attenuate more and lower frequencies less.

The acoustic energy produced by ships and submarines can be due to the following:

1. The propulsion system such as the engine and reduction gears. The noise

of these is typically of narrowband nature because the parts have natural frequencies and harmonics. Friction, on the other hand, can cause wideband noise.

2. Propeller. The rotating blades cause cavitation noise which is of wideband nature.

3. Auxiliary machinery such as pumps and electrical generators. The noise is typically tonal i.e. narrowband.

4. Hydrodynamic effects such as radiated flow noise.

The speed of the vessel also affects the noise emitted by it. [34, pp. 363–365]

The same source [34] gives also representative total spectrums of ships and submarines. Measured close to the source, they start do decay at 100 Hz and have decreased 20 dB until 1 kHz and another 20 dB at 10 kHz.

The signal received at the sensor array is thus dependent on the source signal itself, the absorbtion loss which grows with increasing distance and the ambient noise present in the sea. Measurements with modern commercial ships made in the sea have been published in [35]. They use a container ship, a bulk carrier and a tanker to drive past the measurement system and give spectrograms that visualize the frequencies in the received signal. At the closest point to the measurement system the ships are at a distance of 3 km and at the farthest at 20 km. The energy of all the signals is concentrated at the frequency range under 100 Hz. The container ship and the bulk carrier rise above the noise level in the frequency range 100–1000 Hz only when the distance is 5 km or less and the spectrum of the tanker shows even less energy in that range.

When these numbers are compared to Figure 23 it becomes clear that the problem of detecting close sources is challenging. We would want to use as much information from the high frequencies since the resolution is much better but also accept the fact that most source signals are strong and can only be existant on the low frequency where a lower resolution ability is apparent.

The amount of ambient noise in the ocean somewhat decreases with growing frequency according to [34]. It is clear for us, that mostly the SNR is very low: around 0 dB which means that the information containing signal and the ambient noise are as powerful. It is most important that the noise is spatially white because this is the assumption made in the subspace methods. This might not be completely true in reality: Because the hydrophones are close to each some perturbations caused by the water flow for example can affect all the hydrophones similarly which makes the noise correlated.

7.3 Tracking in real time

The DOA estimation algorithm is meant to be used live in the underwater environment. The sampled signals at the hydrophones can be processed in patches which are in this thesis called *frames*. New information about the locations of the sources is wanted continuously so ideally these frames would be as short as possible. The sources can also move during the time that is needed to aquire data for a single frame which degrades the fit to the DOA model. On the other hand, we need a number of snapshots to improve the estimates for e.g. the covariance matrices. A sensible compromise for the length of one frame is one second which is held fixed in the thesis and means that for $K = 64$ frequency bands we have $P = 8000/(2 \cdot 64) \approx 62$ snapshots.

The visualization of the source locations that an operator can use is a spatial spectrogram where the magnitudes (powers) of sources are shown as a function of time and angle. Such a plot is given in Figure 28, for example. Typically these three dimensional plots are shown directly from above and the intensity of the color describes the magnitude. Most of the DOA methods used in this thesis produce a continuous spectrum. The spectrogram can consist of these continuous curves or – in principle – we can just show the locations of the peaks and leave everything else out. This can produce a much cleaner plot (and is neccessary if the information is utilized for automatic purposes) but on the other hand the shape of the continuous curve can give extra information to the operator about the uncertainty of the estimate. For example, within a single method, a sharper peak suggests a more certain estimate while a wide lobe can be caused by a structured noise source.

8 Results

In this section the wideband DOA estimation methods and the sparse modeling of the sensor data is tested. The methods for wideband DOA estimation are tested with simulated scenarios and with real data while the sparse modeling is applied to real data only. For the DOA estimation methods, the real data is a good indication of robustness – some methods might fail altogether as will be shown later on. Unfortunately, the absolute truth about the routes of the sources, the number of all potential noise sources etc is unknown. Thus, the real data can not be used to measure the resolution ability or accuracy exactly. These scenarios are produced with synthetic data, the generation method of which is decribed in Section 8.1.

The following wideband DOA estimatin methods were implemented in MATLAB and used in the following subsections:

- MVDR-S, Algorithm 4
- MUSIC-S, Algorithm 5
- MVDR-LBC, Algorithms 11 and 12
- CSSM, Algorithm 6
- WAVES, Algorithm 7
- TOPS, Algorithm 8
- OMP-W, Algorithm 10.

The sensor array and processing parameters were described in Section 7. They are gathered here:

- The sensor array is ULA with 0.25 m distance between the sensors.
- There are $N = 18$ sensors.
- The sampling rate is $f_s = 8000\,\text{Hz}$.
- The number of snapshots is $P = 62$.
- The number of total frequency bands is $K = 64$.

The angle grid has a resolution of $\vartheta_{m+1} - \vartheta_m = 0.5^\circ$ if not otherwise mentioned. If the precision is important, a finer grid is used.

8.1 Simulations

To test the methods synthetic data is used. We use two kinds of source signals: white noise and autoregressive (AR) processes. Recalling from Section 7 that higher frequencies are easier to resolve than low frequencies, the white noise, having by definition a flat frequency (amplitude) response, works as a fair starting point to test the methods. On the other hand, in Section 7 it was also discovered that the

source signals typically have more power at lower frequencies. A natural model for simulating such sources is the AR process which is used in Section 8.1.2.

Generating the signals received at each sensor requires a little work. In order to know exactly the signal impinging on each sensor we would need samples that are between those sampled at the first sensor with its natural sampling rate. To get a good approximation of the true signals at each sensor, the source signals are first generated with a sampling rate ten times higher than the rate at each sensor. To achieve this the AR coefficients are zero padded as follows:

$$b_k = \begin{cases} a_n & \text{for } k = 10n, n \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases} \qquad k = 0, \ldots, 10p - 1 \tag{120}$$

i.e. every 10'th coefficient is equal to the 'prototype' filter and other coefficients are zero. As is well known in multirate signal processing, the zero padding produces a new signal having mirrors in the frequency response. To get rid of the extra frequencies the output is filtered with a lowpass FIR filter visualized by Fig. 26. Then the exact value for the time instant required by the sensor is acquired by interpolation using the function `interp1` and the method `pchip` which is a cubic interpolation method in MATLAB.

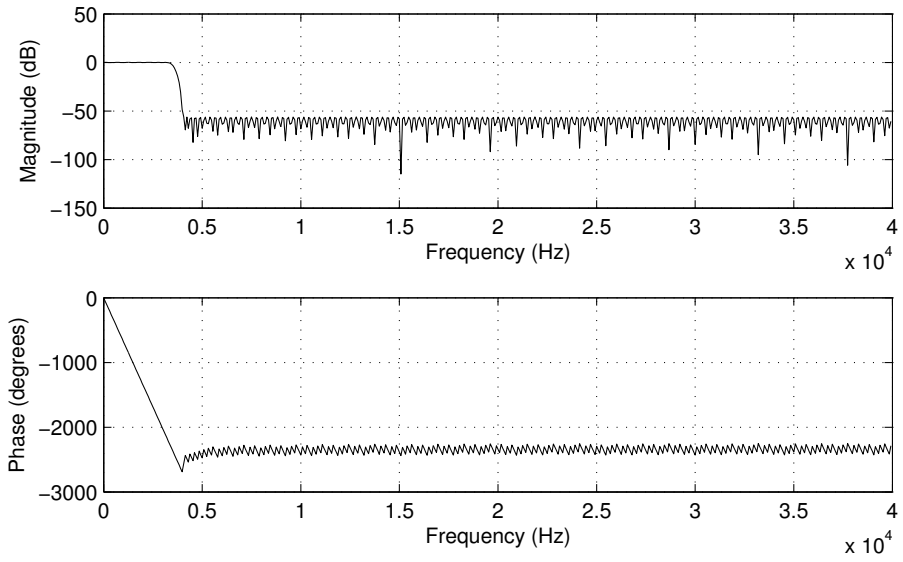


Figure 26: The FIR filter used in the generation of the synthetized signals.

8.1.1 Moving white noise sources

Using white noise as the source signal seems like an objective choice in order to truly see differences between the DOA estimation methods. The frequency spectrum is flat so no frequency bands become more important than others.

The scenario, illustrated in Figure 27, consists of three sources. The first moves from 50° to 54°, the second from 60° to 58° and the third stays constantly at 90° during the 70 second experiment. One frame corresponds to processing one second.

The wideband spectra $\mathcal{S}$ produced by the methods are illustrated for two selections of the subband indices used and for three different SNR. For the case where all the usable subbands are used i.e. $\mathcal{K} = \{6, 7, \ldots, 46\}$ the results are depicted in Figure 28 for the SNR of 5 dB, in Figure 29 for 0 dB and in Figure 30 for -5 dB. The source at 90° is cropped out since it is well resolved by all the methods in all the cases.

In the case where only high frequency subbands are used i.e. $\mathcal{K} = \{30, 31, \ldots, 46\}$ the results are depicted in Figure 31 for the SNR of 5 dB, in Figure 32 for 0 dB and in Figure 33 for -5 dB.

Several observations can be made from the figures:

- The methods can be divided into two groups: the promising ones MVDR-S, MUSIC-S, MVDR-LBC, CSSM and WAVES and the weak ones: TOPS and OMP-W. In the forthcoming simulations, TOPS and OMP-W are not used because of their poor performance.
- The poor performance of TOPS especially when $\mathcal{K}$ consisted of more indices is mostly due to the bad selection of $\boldsymbol{U}_0$ (see Section 4.3.2). In lack of better knowledge it was chosen to correspond to the first frequency index which is clearly a bad decision considering the discussion about beamwidth made in Section 7. When $\mathcal{K}$ consisted of the high frequencies only, TOPS has a better chance of working well and it does. However, it does not beat MUSIC-S, CSSM or WAVES.
- OMP-W has poor resolution because of the beamformer-like nature of it discussed in Section 3.4.1. It was the only method to which the true number of sources was given. Only the peak shown at 55 degrees is truly caused by the sources and the other one is just due to going one iteration too far. More understanding for this selection can be gained by looking at Figure 11.
- Lowering the SNR seems to weaken all the methods quite evenly except that WAVES seems to work better than CSSM when the SNR is very low at -5 dB.
- MVDR-LBC much improves the visibility of the sources compared to MVDR-S. The contradicting information of the subbands is seen as extra lobes appearing in between the sources.
- CSSM and WAVES become biased when the low frequencies are present ($\mathcal{K}$ has more bands). Otherwise, they seem to offer the best resolution.

In addition it was noticed that the information theoretic criteria used in the selection for the number of sources generally work well with the simulated data. An example of the criterion is shown in Figure 34. Also, during these experiments it was discovered that if the peak locations of MVDR-S is not very symmetric, CSSM and WAVES using it as the initialization method can produce a strange spectrum as is depicted in Figure 35. This implies that they are quite sensitive to the initialization.

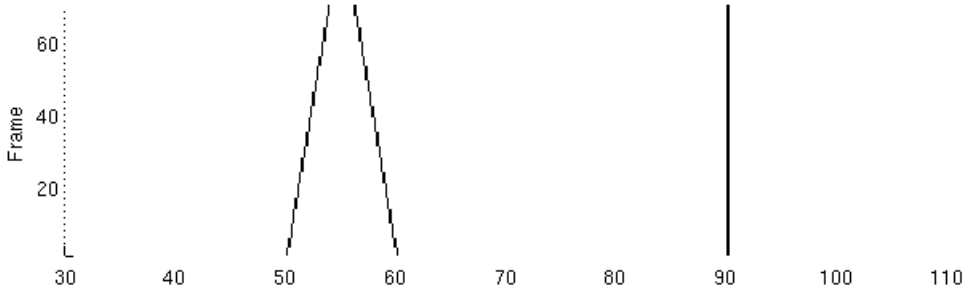
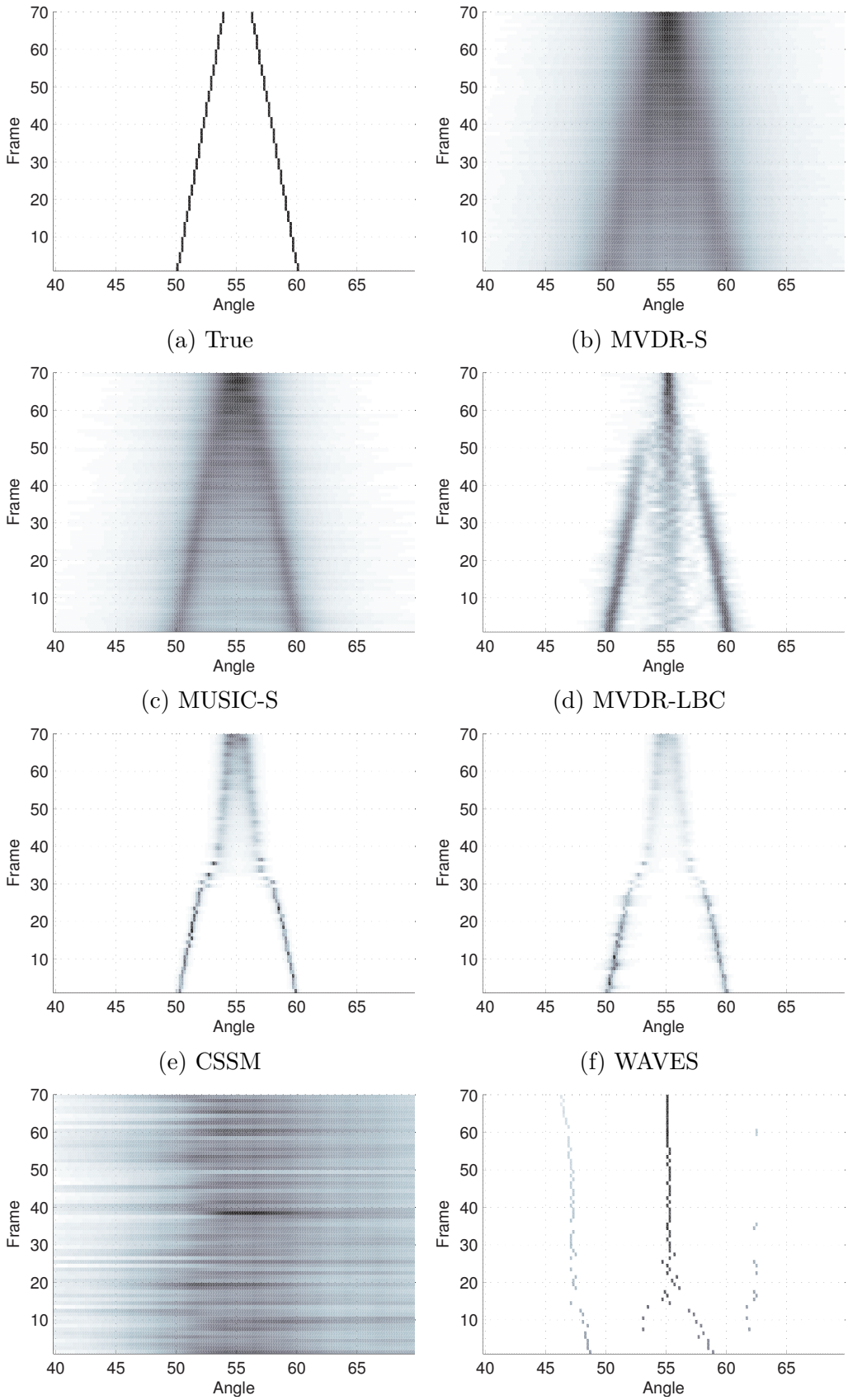


Figure 27: True scenario.



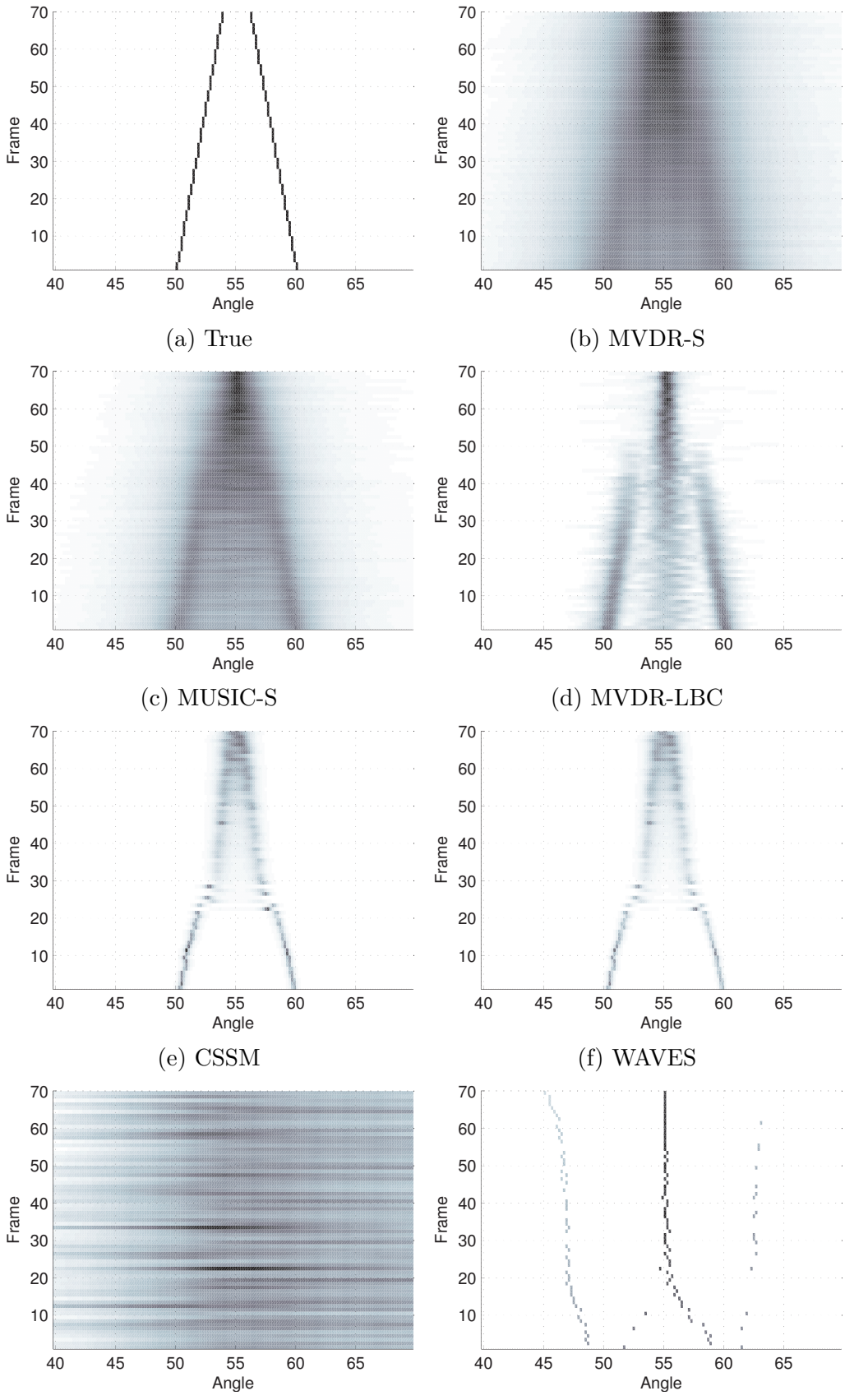
(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

(g) TOPS (h) OMP-W

Figure 28: SNR is 5 dB. Frequency bands from 6 to 46 with white noise source.



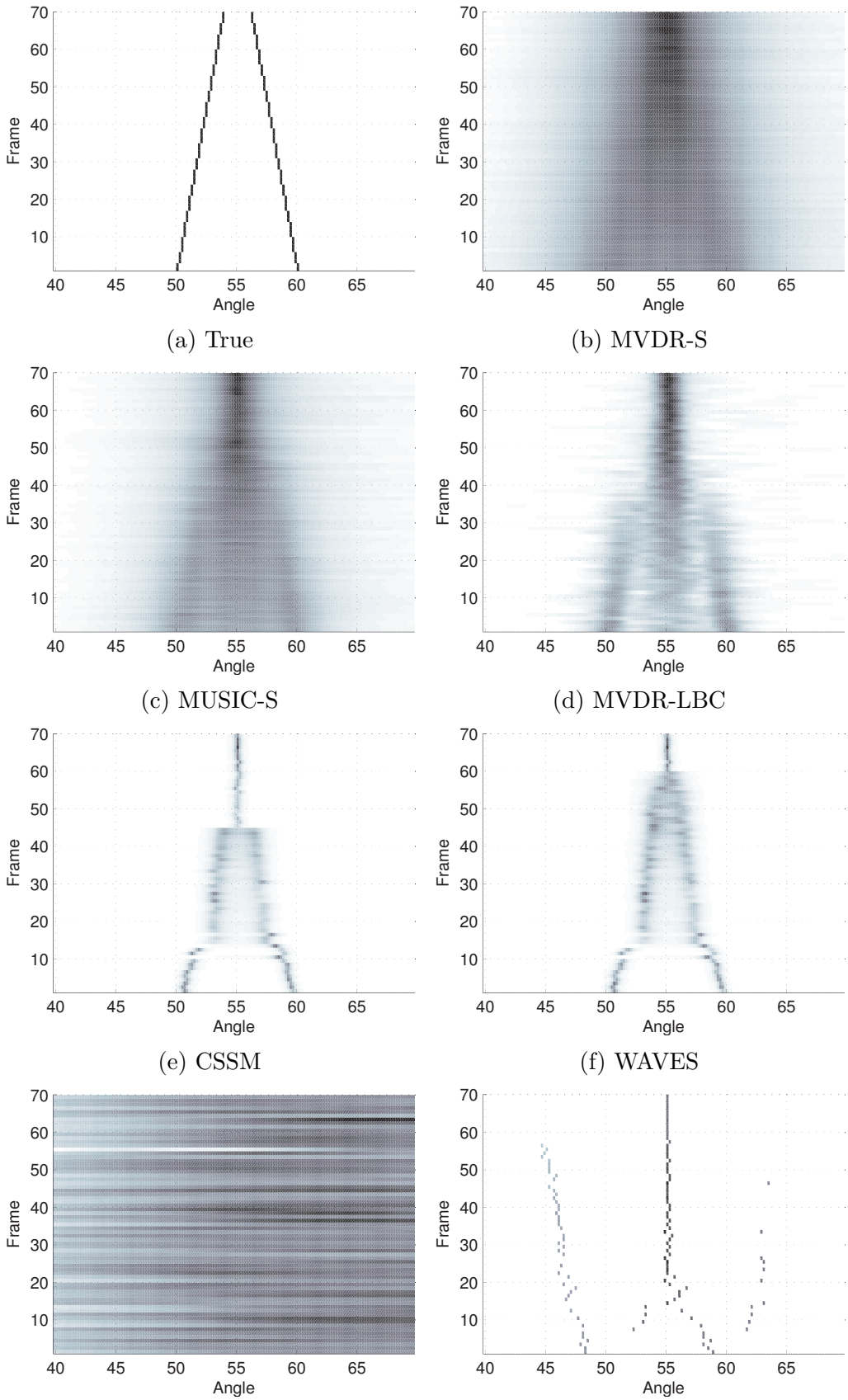
(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

(g) TOPS (h) OMP-W

Figure 29: SNR is 0 dB. Frequency bands from 6 to 46 with white noise source.



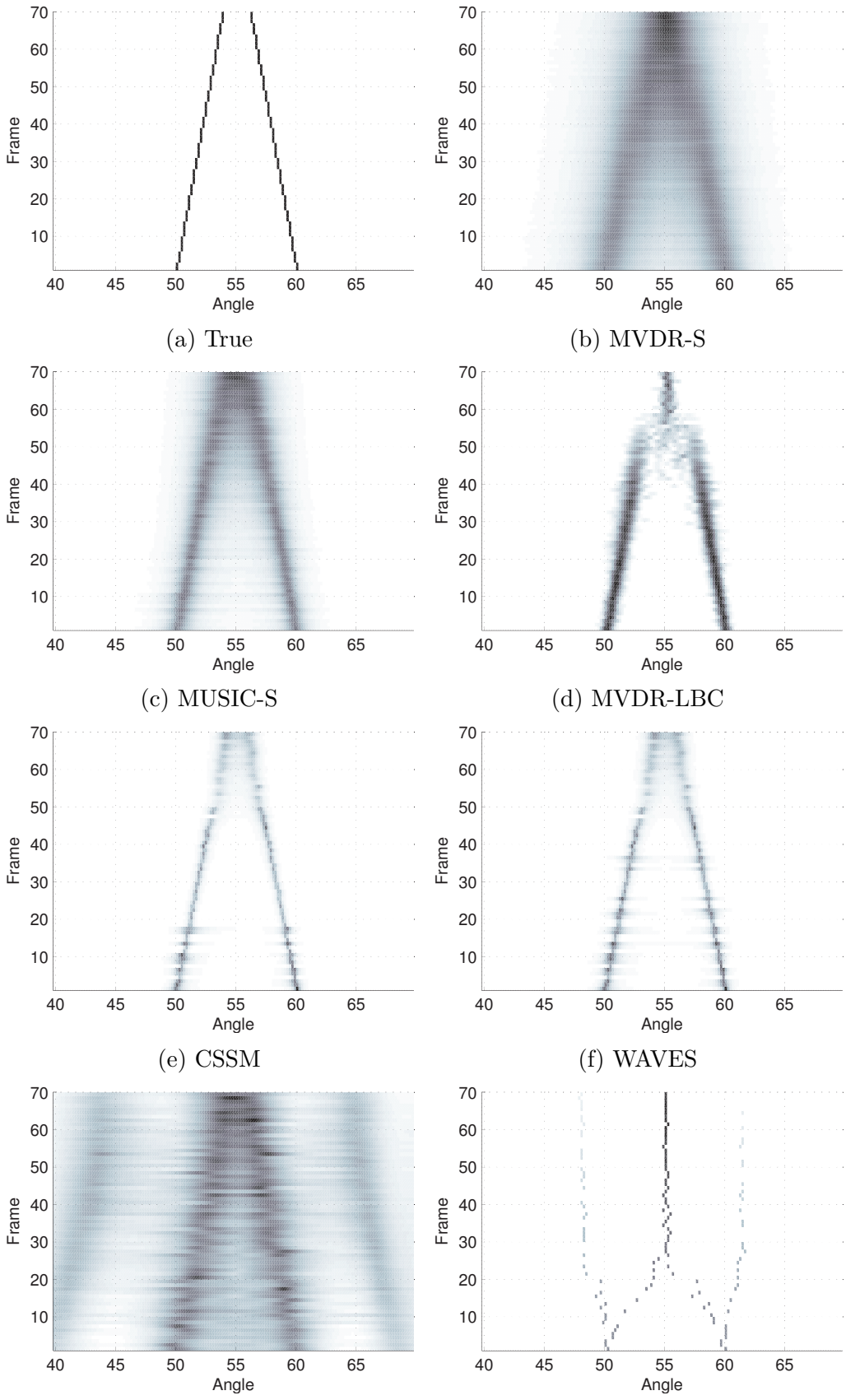
(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

(g) TOPS (h) OMP-W

Figure 30: SNR is -5 dB. Frequency bands from 6 to 46 with white noise source.



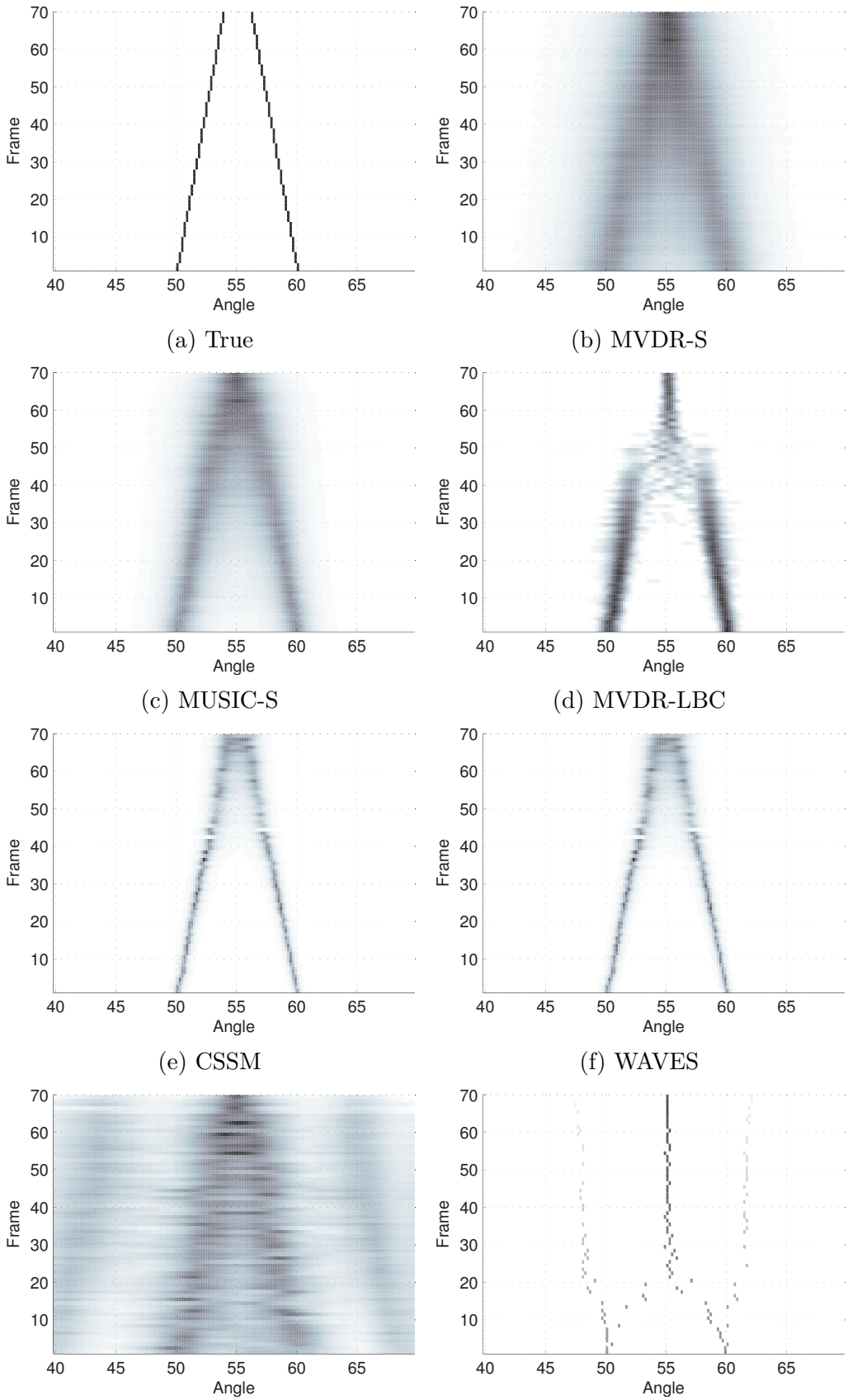
(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

(g) TOPS (h) OMP-W

Figure 31: SNR is 5 dB. Frequency bands from 30 to 46 with white noise source.



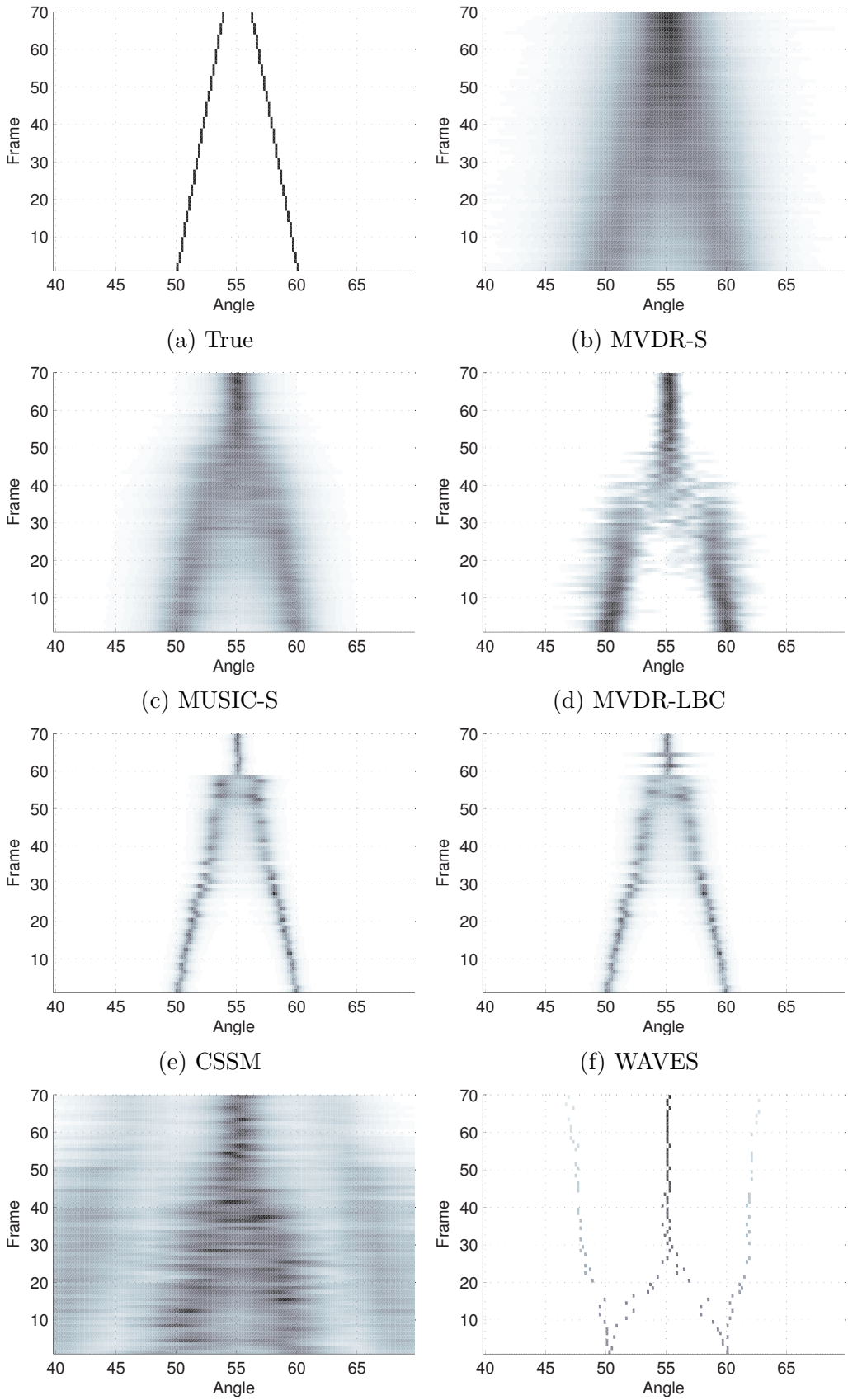
(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

(g) TOPS (h) OMP-W

Figure 32: SNR is 0 dB. Frequency bands from 30 to 46 with white noise source.



(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

(g) TOPS (h) OMP-W

Figure 33: SNR is -5 dB. Frequency bands from 30 to 46 with white noise source.

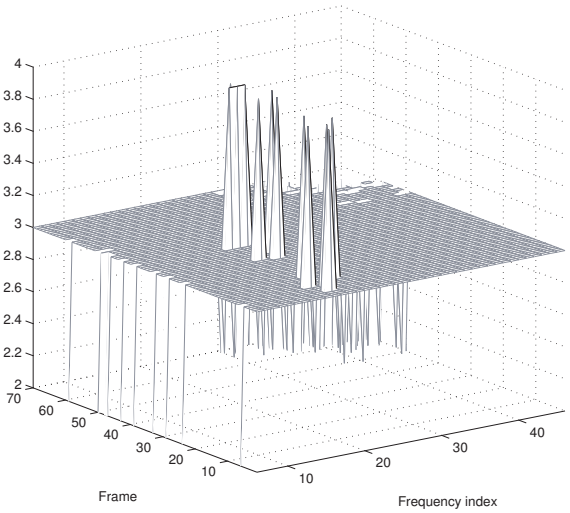


Figure 34: The selected number of sources for MUSIC-S. The correct choice is 3. It can be seen that only in the very last frames the choice becomes systematically wrong.

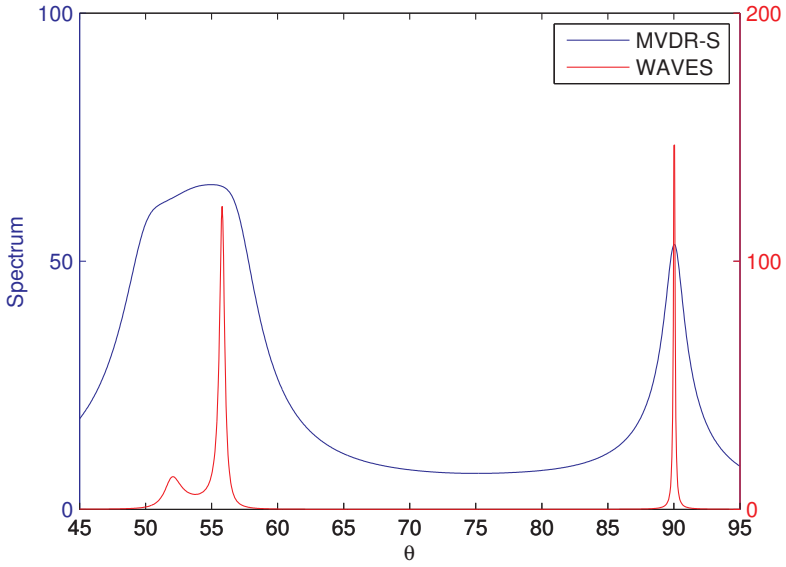


Figure 35: The coherent methods are sensitive to the inital angle estimates. Here one of the sources is highly attenuated because MVDR-S (giving the initial value) peaks only at the second source.

8.1.2 Moving AR sources

In this section the methods are tested in a more realistic scenario using the AR sources whose amplitude responses are depicted in Figure 36. The angle grid has a resolution of 0.2°.

The first experiment depicted in Figure 38 consists of three sources: the first of type source 2 (see fig 36) moves from 50° to 54°, the second is of type 1 and moves from

60° to 58° and the third is of type 1 and stays constantly at 80°. The SNR of the source signals at the sensors is 5 dB for each. Frequency bands from 15 to 46 are used. The smallest frequency corresponding to frequency 875 Hz seems quite high considering the discussion in Section 7 about the real signals but we do not want to distract the methods too much with the contradicting low frequency information.

The methods OMP-W and TOPS are left out since they do not provide comparable performance as can be seen from the previous experiments. Instead, we add another method which is CSSM but such that instead of MVDR-S it uses MUSIC-S as the method for acquiring the pre-estimates for the DOAs. This is because from other experiments it has been discovered that the initialization is important for the coherent methods.

From Figure 38 it can be seen that MVDR-S gives good indication about the sources: they seem equal in power and the wide lobe around 55° implies that there are more than a single source. MVDR-LBC, on the other hand is able to visualize the locations of the true sources much better and still retain the equality of the powers. MUSIC-S can see two sources in the area around 55 degrees further than either of the MVDR methods.

CSSM and WAVES in their original form give biased estimates and the magnitudes depicted by the intensity of the lines varies much more than those of the incoherent methods. At frames 10 – 20 they lose one of the sources altogether because the initialization method (MVDR-S) does not show a peak in there. Later on, there are two peaks visible around the center angle of the true sources, though. This implies that the coherent methods indeed work only on a narrow neighbourhood of the initalization angles. In Figure 38c the initialization was done with MUSIC-S instead and this definitely improves the performance in frames 0 – 40. On the other hand, plain MUSIC-S is able to distinguish the sources until this point.

The Figure 40 shows a similar situation to 38 except that there are two additional sources. The fourth source is of type 2 (see fig. 36) and moves from 100° to 105° and the fifth source is of type 1 and moves from 120° to 110°.

The performance of the incoherent methods remains similar to Figure 38 as is expected: Since the new sources are far enough they should not have a big impact on the resolving of the sources. In principle, the additional sources can affect the coherent methods because there are more angles in which to cohere. Indeed, it can be noticed that even the resolution ability of CSSM using MUSIC-S as initialization is degraded slightly. The sources around 55° become sometimes indistinguishable between frames 30 and 40.

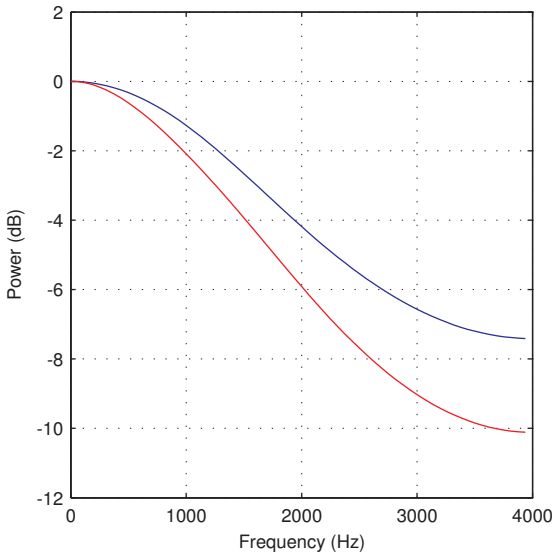
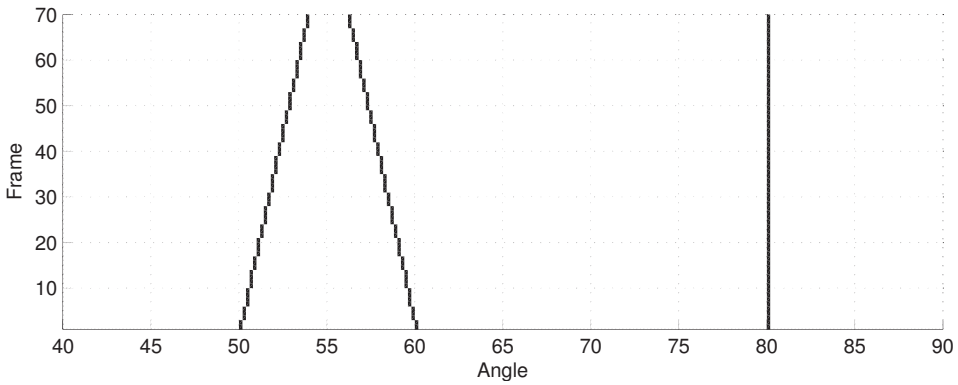
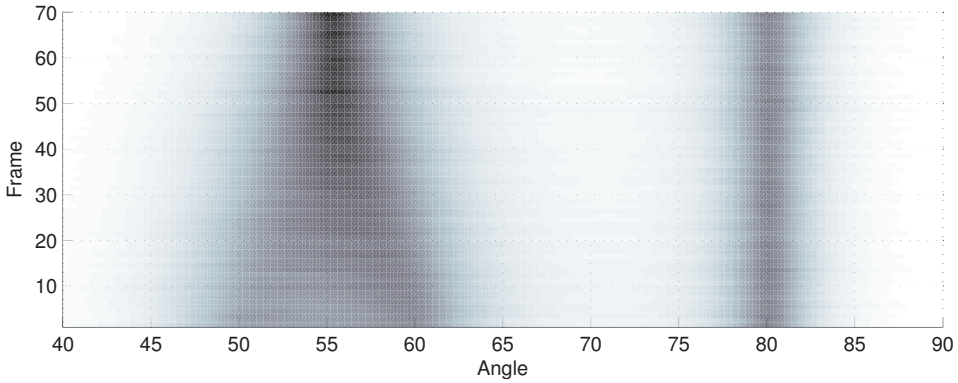


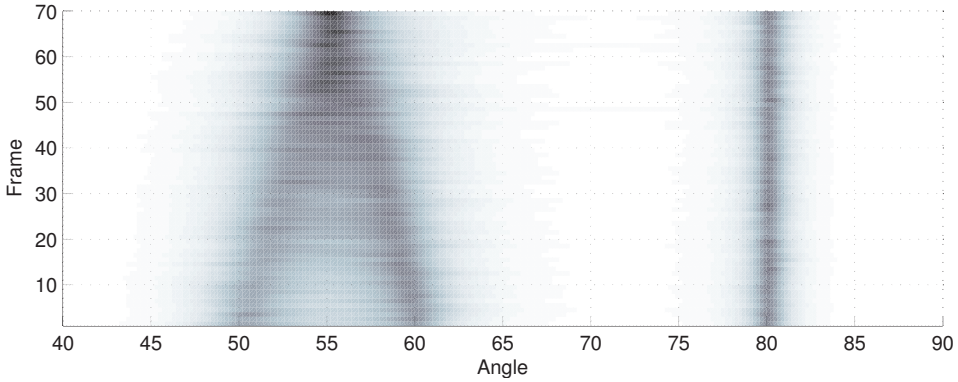
Figure 36: The frequency spectra of the source signals. The blue curve corresponds to source 1 and the red to source 2.



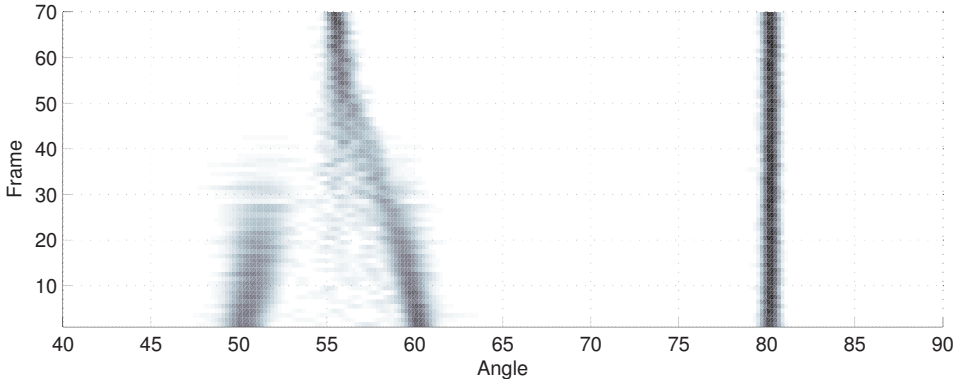
(a) True



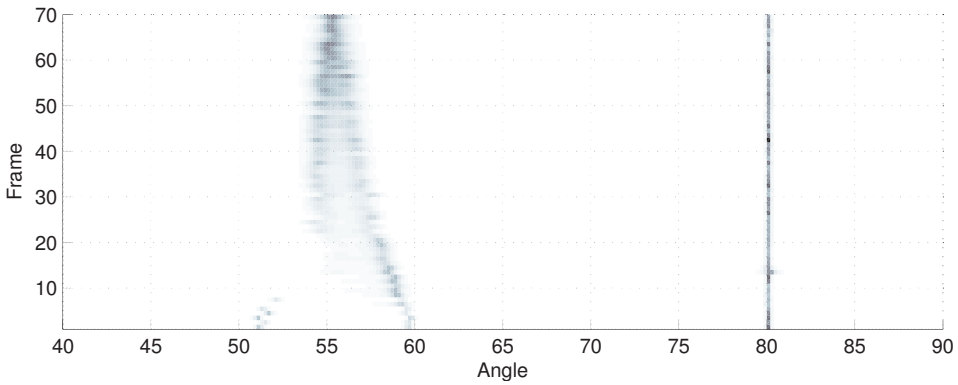
(b) MVDR-S



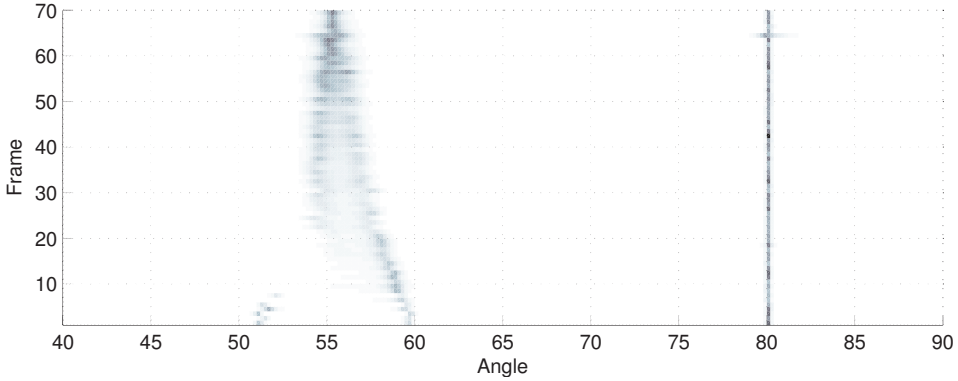
(c) MUSIC-S



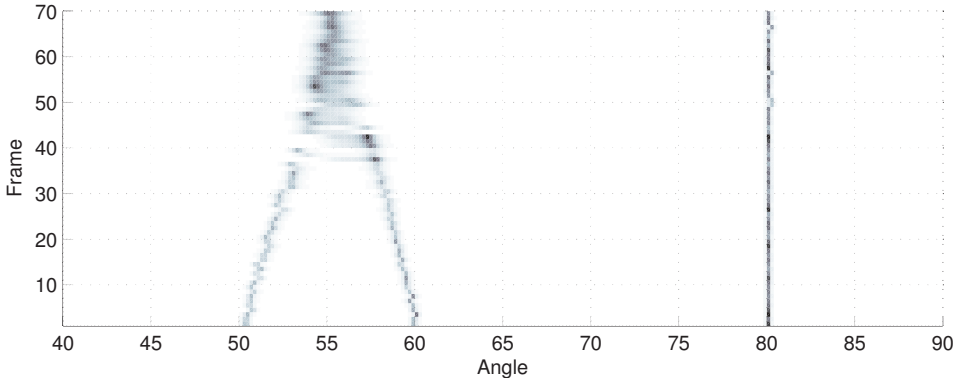
(d) MVDR-LBC



(a) CSSM

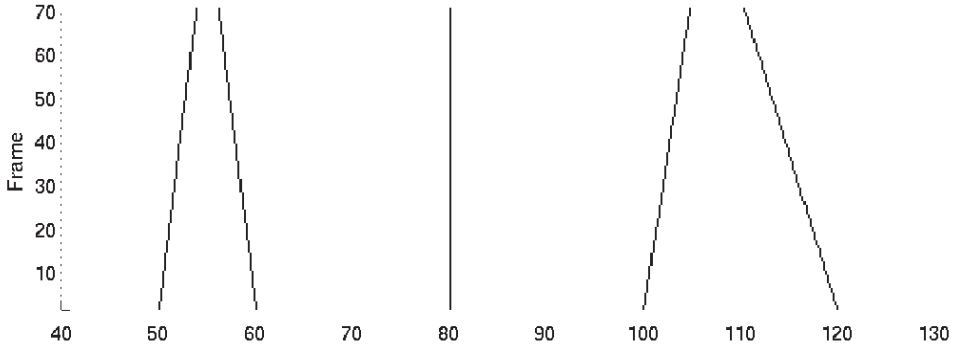


(b) WAVES

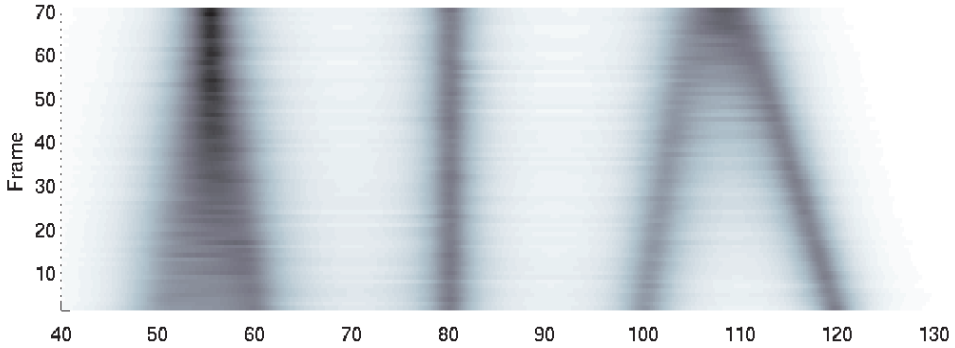


(c) CSSM with MUSIC as initialization.

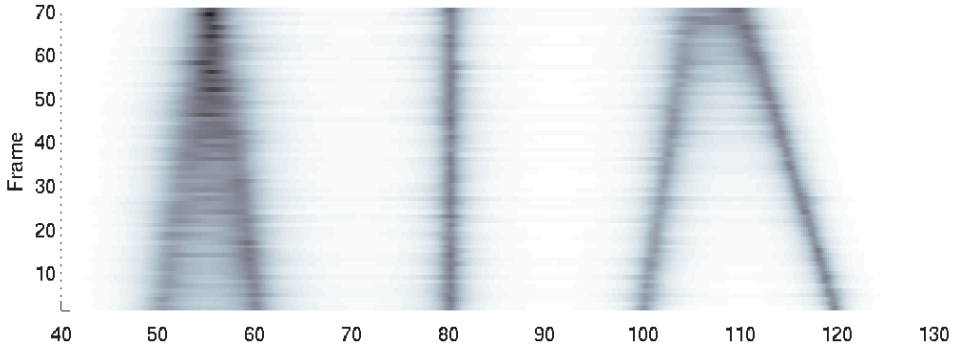
Figure 38: The spectrograms obtained with the methods for 3 sources.



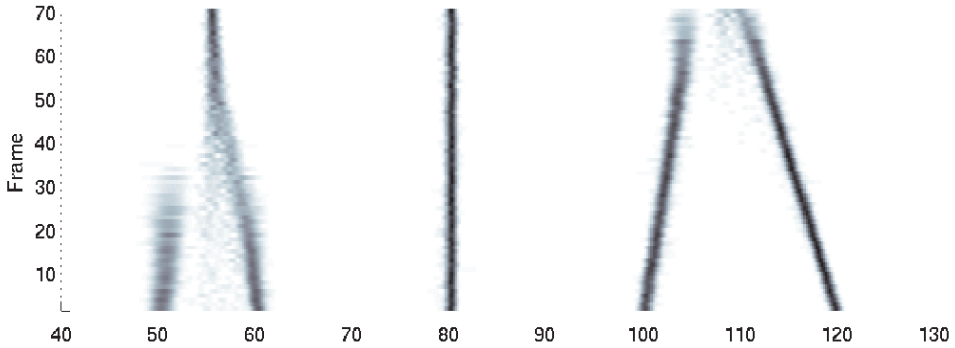
(a) True



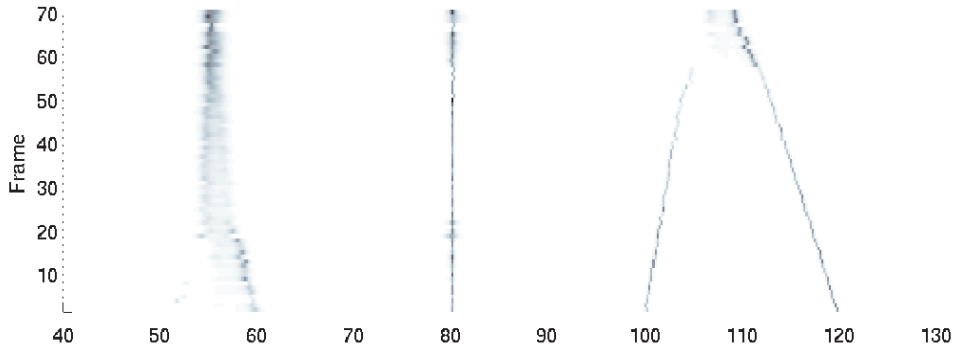
(b) MVDR-S



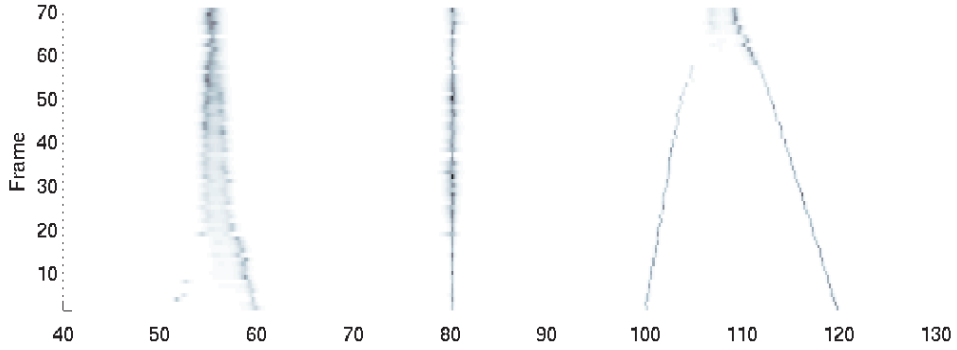
(c) MUSIC-S



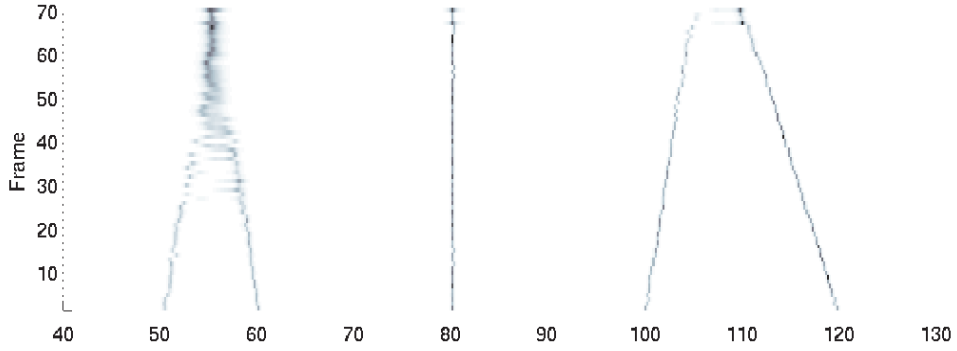
(d) MVDR-LBC



(a) CSSM



(b) WAVES



(c) CSSM with MUSIC as initialization.

Figure 40: The spectrograms obtained with the methods for 5 sources.

8.1.3 Time varying power of the sources

In the real environment the source signals naturally have different powers at the sensors because they are at different distances and have different acoustic noise sources. This kind of variation is prone to have an impact also on the DOA estimation: a strong source can mask a weaker one. This section simulates the variation of source powers and shows spectrograms produced with the different DOA estimation methods.

The simulation is done as follows: All the sources stay put. The first source is an AR process of type 1 (see 36) at 50°, the second an AR process of type 2 at 59° and the third an AR process of type 2 at 90°. The SNR of the 1. source at 50° is held fixed while the SNR of the 2. source at 59° is varied with time. The source at 90° also has fixed SNR at 0 dB and is left out of the result figures because no interesting phenomena was noticed related to it. We process 100 seconds which corresponds to 100 frames. The SNR of the sources 1 and 2 is depicted in Figure 41.

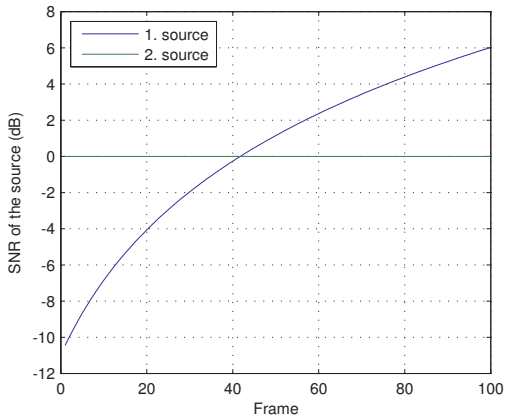
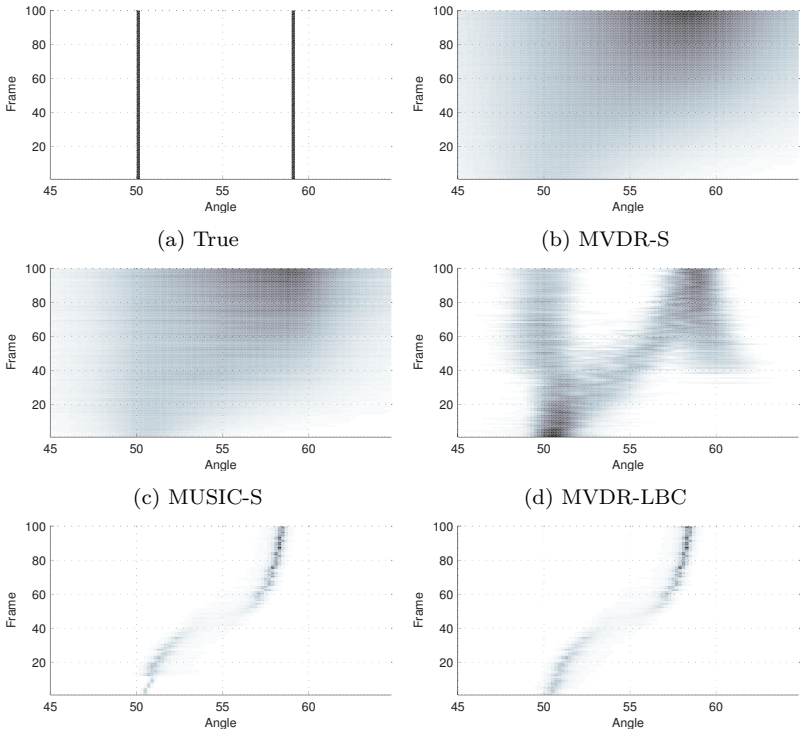


Figure 41: The SNR of the source signals at the sensors with respect to the frame index.

The resulting wideband spectra are depicted in Figure 42 for $\mathcal{K} = \{6, 7, \ldots, 46\}$ and in 43 for $\mathcal{K} = \{30, 31, \ldots, 46\}$. Notice that the figure describing the truth only illustrates the position of the sources, not the magnitude. The task is simple for all the methods when only the high frequencies are taken into account. It is interesting to notice that when the low frequencies are present, another strong close source biases the estimation for all other methods than MVDR-LBC such that they only appear as one moving source.

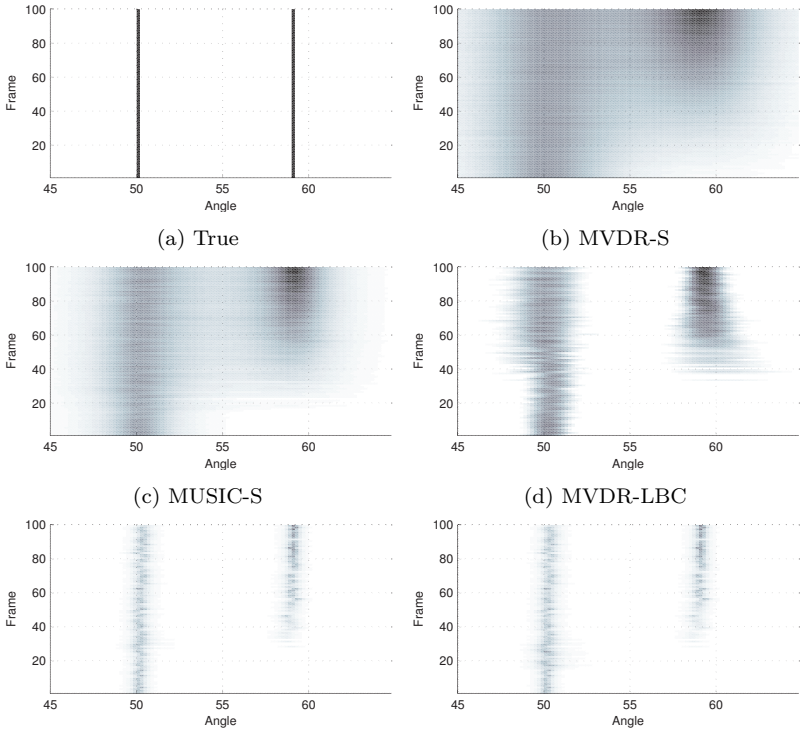


(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

Figure 42: The wideband spectra when the powers of the sources are changing. Subbands from 6 to 46 are used.



(a) True (b) MVDR-S

(c) MUSIC-S (d) MVDR-LBC

(e) CSSM (f) WAVES

Figure 43: The wideband spectra when the powers of the sources are changing. Subbands from 30 to 46 are used.

8.1.4 Statistical evaluation of resolution

In order to compare the methods statistically, a test where the sources are held fixed for a number of noise realizations is conducted. All the sources stay fixed for $J_i = 200$ realizations. The first source is at 50°, the second at $50° + \Delta\theta$ and the third at 90°. The source signals are white noise with equal power. This process is repeated for $\Delta\theta = 4, 5, \ldots, 10$ and for several values of SNR, namely $-5, -4, \ldots, 5$. The angle grid has a resolution of 0.05°. A single test like this is visualized as a spectrogram in Figure 44. Only the peaks are shown because in this part they are of interest. The term 'Frame' in the figure could also in this case be called 'realization'.

Two types of information is extracted from the tests: the probability of resolution and the accuracy of DOA estimates. This section covers the first and Section 8.1.5 the second.

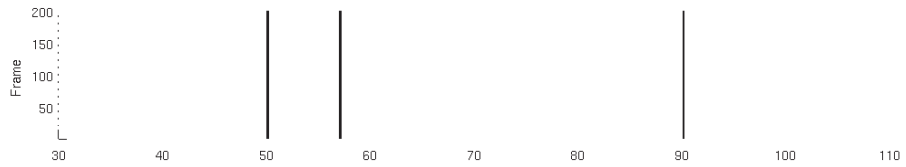
The process of computing the probability of resolution is the following: For each realization:

1. Detect the local peaks.
2. For each true source location, find the peak that is closest to the source.
3. If each peak corresponding to a source is different to the others, the realization is considered resolved.

In the end, the number of successful realizations J_s is calculated and the probability of resolution is defined as $P_s = \frac{J_s}{J_i} \cdot 100\,\%$. Notice that this allows for extra, false, peaks.

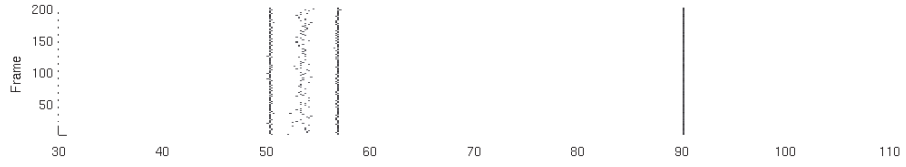
When it comes to the subbands used, the situations are split again into two cases: In the first, most of the frequency content is used: the subbands corresponding to indices from 6 to 46 are used. The results are visualized in Figure 45. In the second, only the high frequencies are taken into account: subbands from 30 to 46. The results are visualized in Figure 46.

The results can be interpreted such that WAVES and CSSM share the same behaviour. They prove out to be the best ones. In both cases, MVDR-LBC is much better than MVDR-S. In the low frequency situation MVDR-LBC seems even better than MUSIC-S because for MUSIC-S the wide lobes appearing in the low frequencies mask the sharper lobes appearing in the high frequencies. It should, however, be remembered that the extra peaks appearing in middle for MVDR-LBC do not show in these results. When only the high frequency bands are taken into account, MUSIC-S and MVDR-LBC behave similarly.



(a) True

(b) MVDR-S

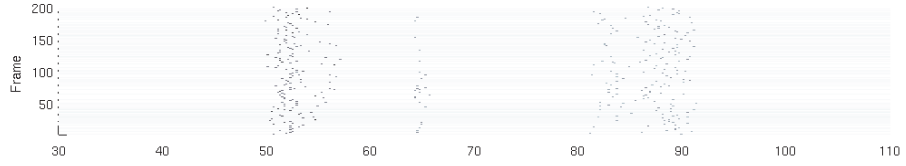


(c) MUSIC-S

(d) MVDR-LBC

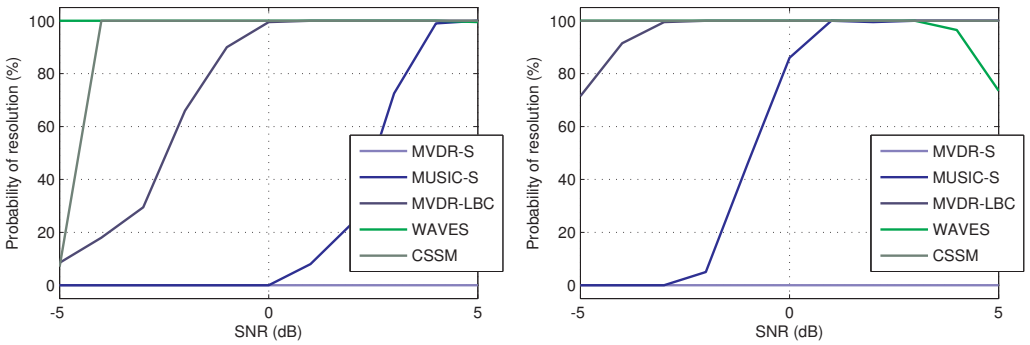
(e) CSSM

(f) WAVES

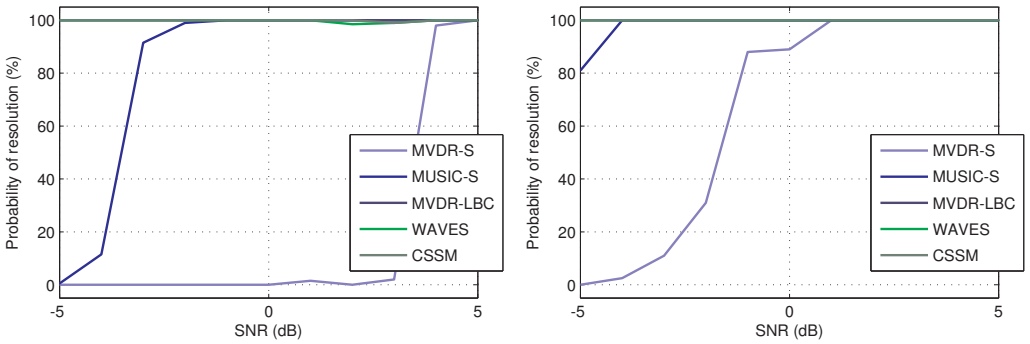


(g) TOPS

Figure 44: The realizations for three sources at 50°, 57° and 90°. SNR is 0 dB.

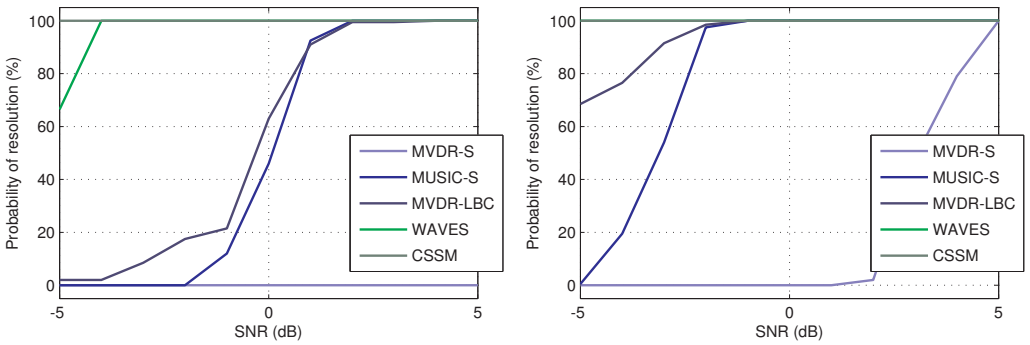


(a) Separation 5°. (b) Separation 6°.

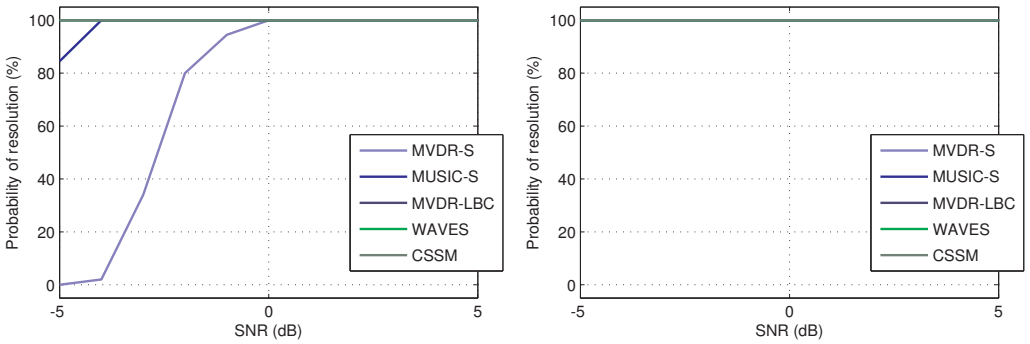


(c) Separation 7°. (d) Separation 8°.

Figure 45: The probability of resolution for different $\Delta\theta$. Subbands from 6 to 46 are used.



(a) Separation 4°. (b) Separation 5°.



(c) Separation 6°. (d) Separation 7°.

Figure 46: The probability of resolution for different $\Delta\theta$. Subbands from 30 to 46 are used.

8.1.5 Statistical evaluation of accuracy

In a similar way to the previous section, the accuracy results for several values of separations $\Delta\theta$ and for the two different subband choices are given. We only give results for the two close sources. The lonely source does not suffer from bias as is depicted by an example in Figure 47. The bias for the close sources is reported in figures 48, 49 and 50 and the standard deviation in figures 51, 52 and 53 for the selection of using subbands from 6 to 46. It can be concluded that MVDR-LBC is very good at compensating for the bias and MUSIC is naturally quite good. CSSM and WAVES behave very similarly and with MVDR-S they suffer from bias. The standard deviation describes how consistent the estimates are. All the methods seem to have small standard deviations until the point of not being able to resolve the sources anymore. It is interesting, however, to note that when the initialization method for CSSM and WAVES, MVDR-S, starts to fail, it causes a rise in the standard deviation of the methods.

The bias is reported in figures 54, 55 and 56 and the standard deviation in figures 57, 58 and 59 for the selection of using subbands from 30 to 46. As is expected, all the methods cope well in this high frequency scenario.

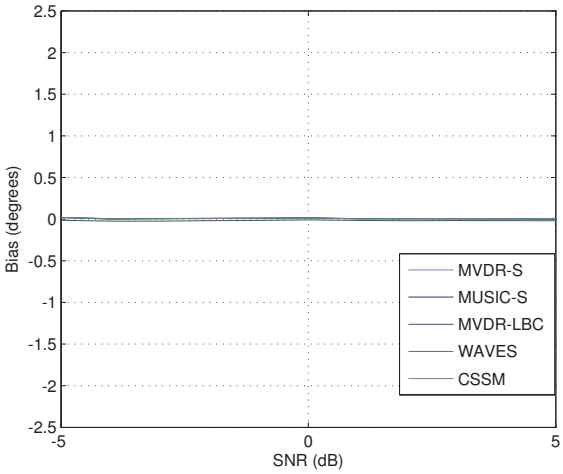
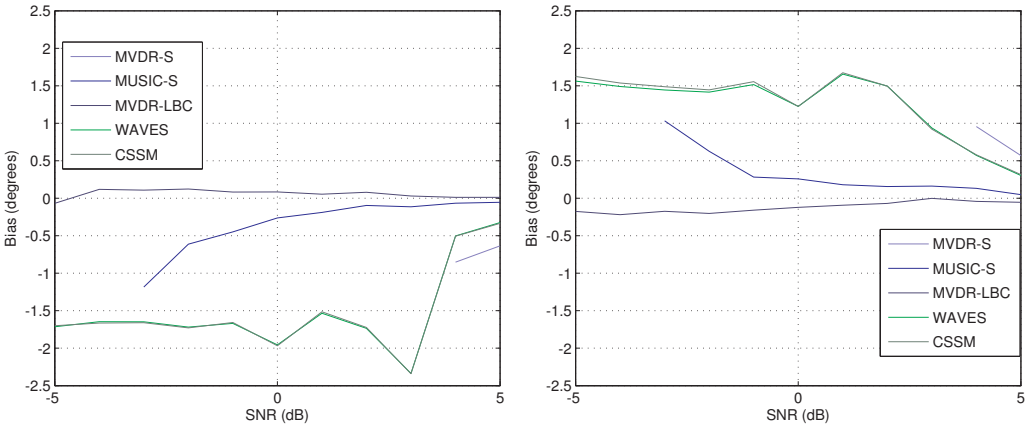
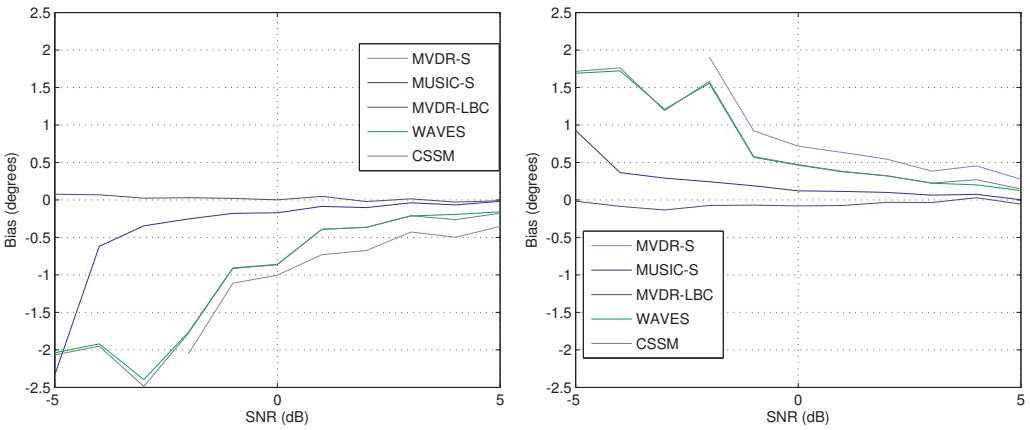


Figure 47: The lonely source at 90° does not suffer from bias.



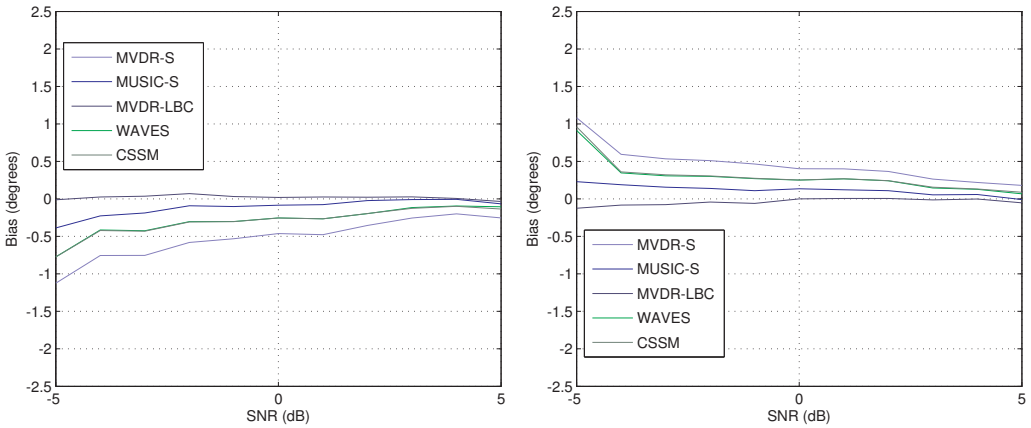
(a) Source at 50°. (b) Source at 57°.

Figure 48: The bias for the separation 7° using subbands from 6 to 46.



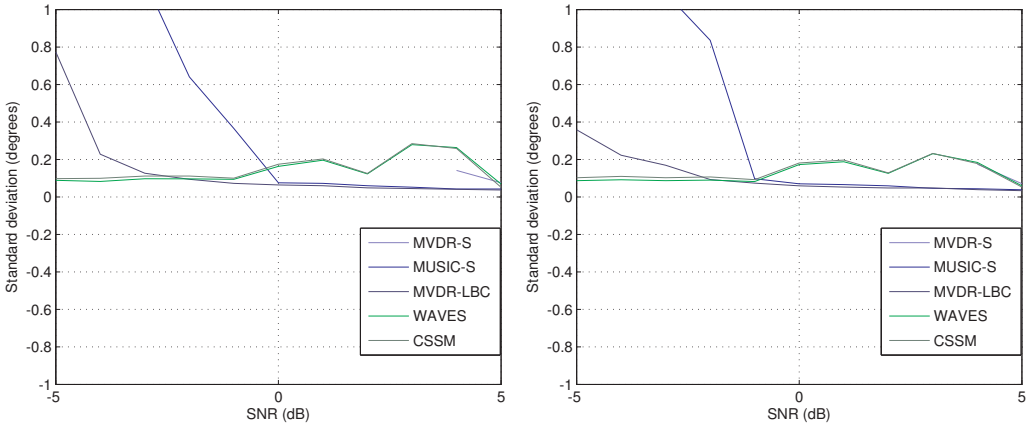
(a) Source at 50°. (b) Source at 58°.

Figure 49: The bias for the separation 8° using subbands from 6 to 46.



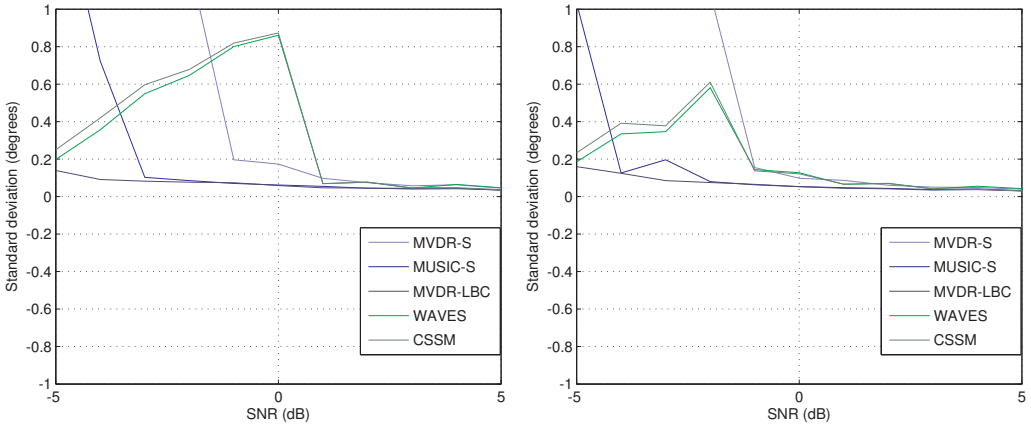
(a) Source at 50°. (b) Source at 59°.

Figure 50: The bias for the separation 9° using subbands from 6 to 46.



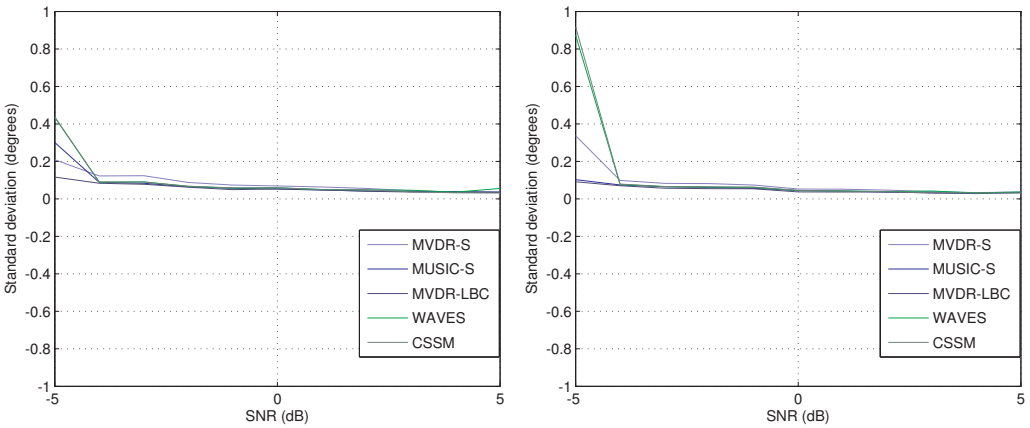
(a) Source at 50°. (b) Source at 57°.

Figure 51: The standard deviation for the separation 7° using subbands from 6 to 46.



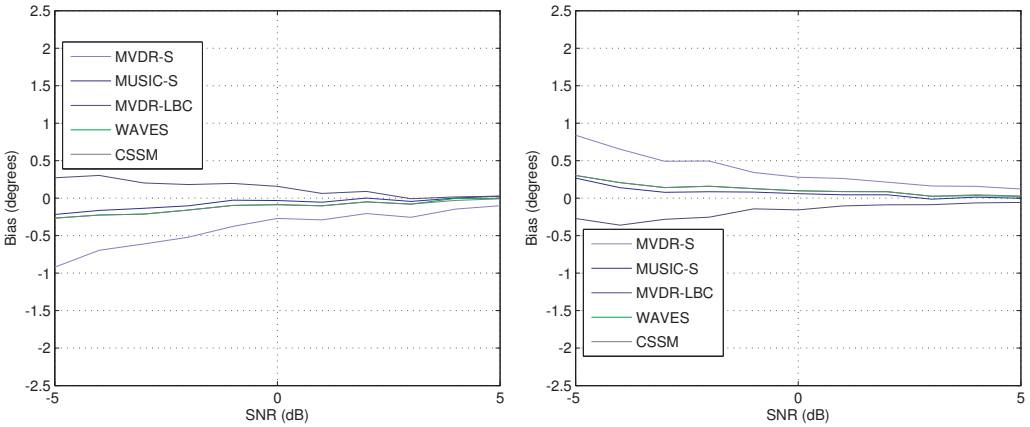
(a) Source at 50°. (b) Source at 58°.

Figure 52: The standard deviation for the separation 8° using subbands from 6 to 46.



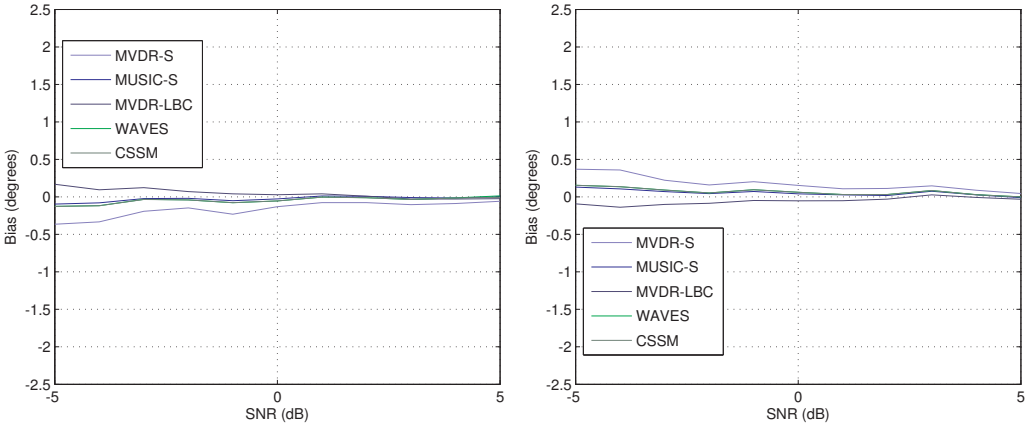
(a) Source at 50°. (b) Source at 59°.

Figure 53: The standard deviation for the separation 9° using subbands from 6 to 46.



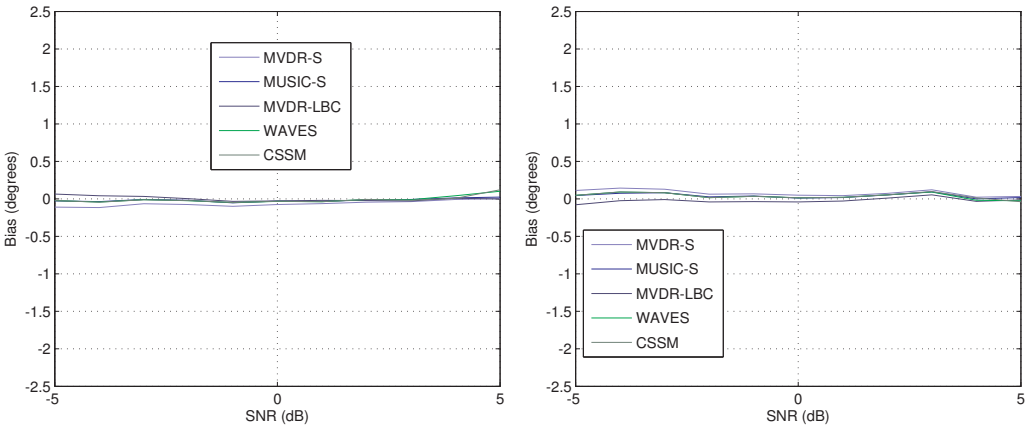
(a) Source at 50°. (b) Source at 57°.

Figure 54: The bias for the separation 7° using subbands from 30 to 46.



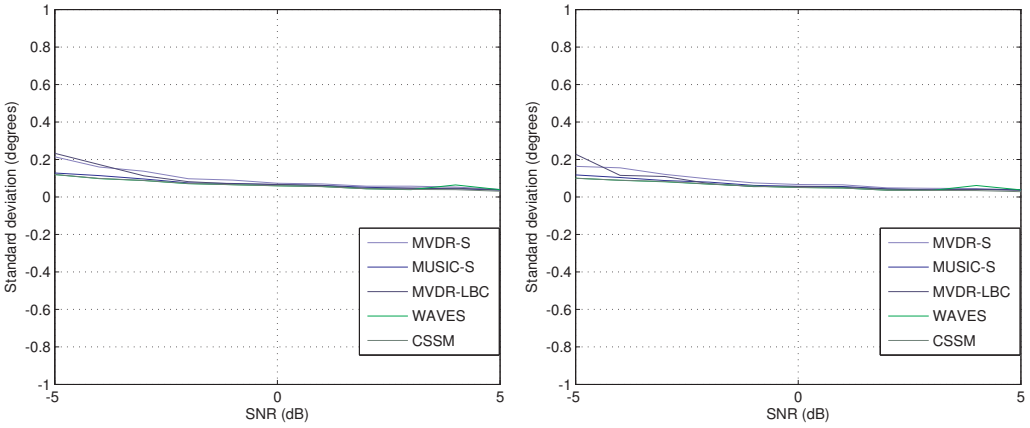
(a) Source at 50°. (b) Source at 58°.

Figure 55: The bias for the separation 8° using subbands from 30 to 46.



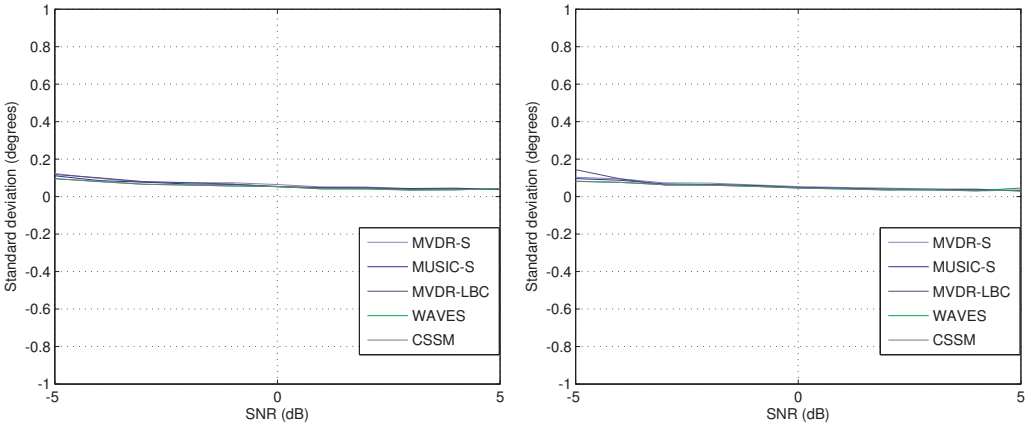
(a) Source at 50°. (b) Source at 59°.

Figure 56: The bias for the separation 9° using subbands from 30 to 46.



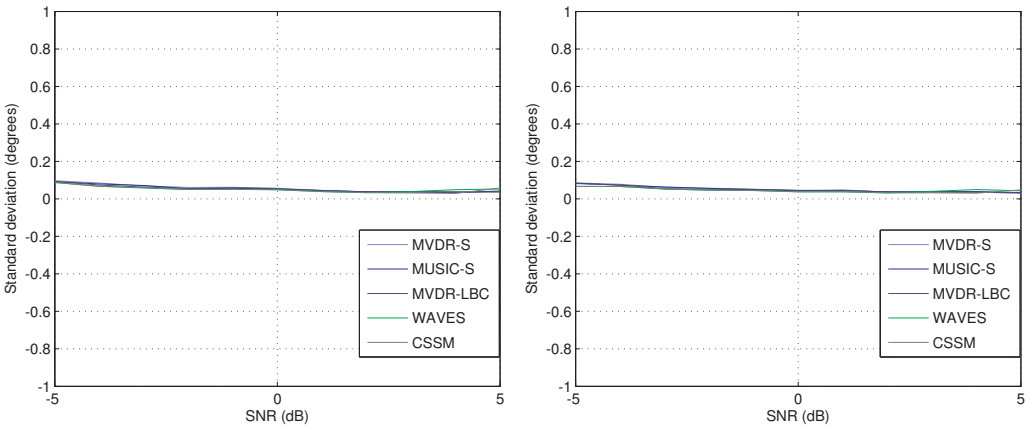
(a) Source at 50°. (b) Source at 57°.

Figure 57: The standard deviation for the separation 7° using subbands from 30 to 46.



(a) Source at 50°. (b) Source at 58°.

Figure 58: The standard deviation for the separation 8° using subbands from 30 to 46.



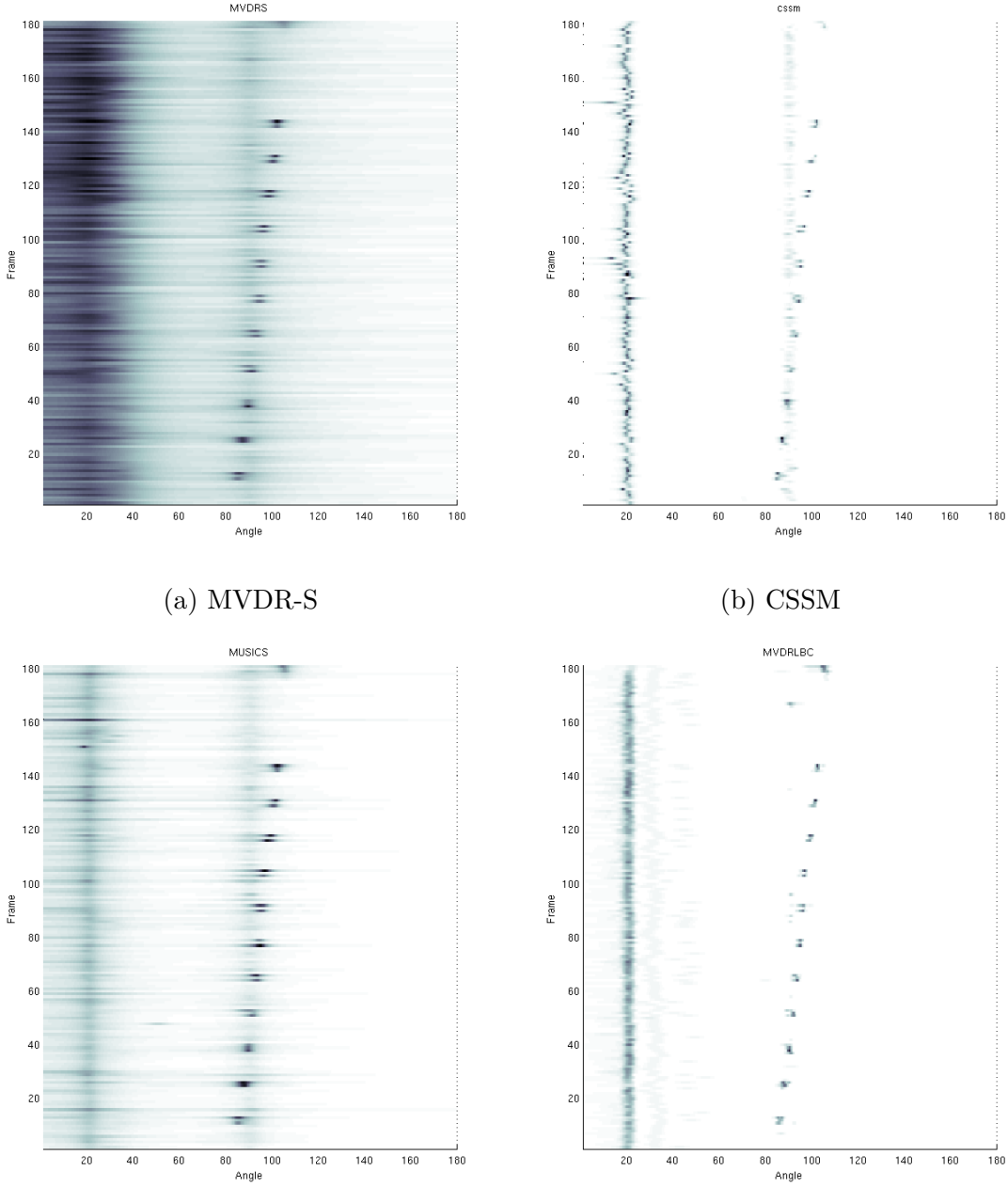
(a) Source at 50°. (b) Source at 59°.

Figure 59: The standard deviation for the separation 9° using subbands from 30 to 46.

8.2 Real data

Real data recorded in the sea is processed. We show two examples containing the wideband spectra of 180 consecutive frames. In the first example shown in Figure 60 a source that sends short impulses is visible. All the methods can see this source well. It seems that MVDR-LBC shows a very consistent line at 20° while some of the others show more variation from frame to frame.

The second example is depicted in Figure 61. In this case, it is discovered that the selection for the number of sources in MUSIC-S is not very successful which is why it is restricted to a maximum of $\lfloor N/2 \rfloor$. The spectrogram without this restriction is shown in Figure 63 and the selected criteria are shown in Figure 62. From Figure 61 it can be said that MVDR-LBC and MUSIC-S with the restriction offer the best visibility of the sources. In CSSM the strong source is very emphasized.



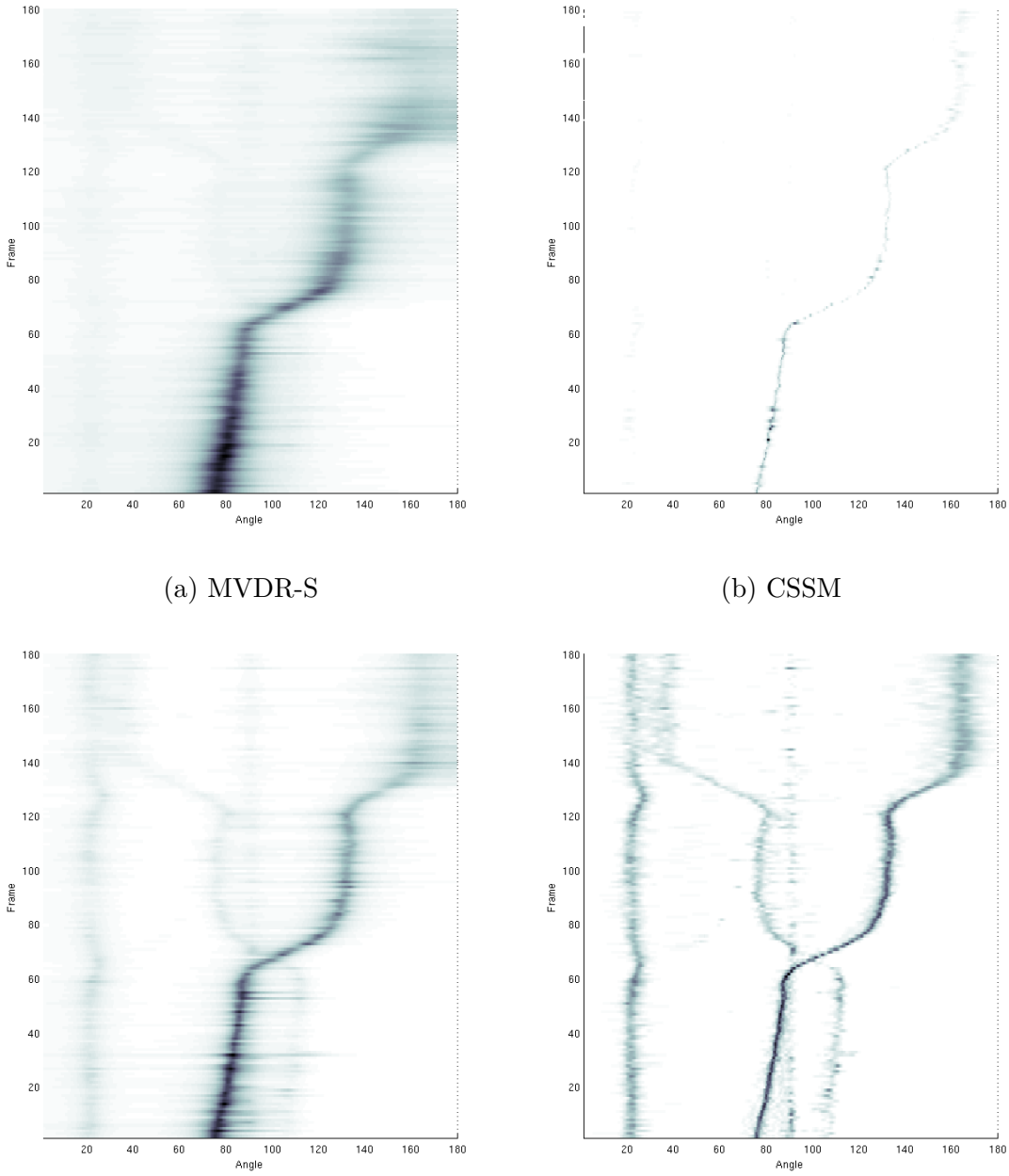
(a) MVDR-S

(b) CSSM

(c) MUSIC-S

(d) MVDR-LBC

Figure 60: Three minutes of real data processed with different methods.



(a) MVDR-S

(b) CSSM

(c) MUSIC-S

(d) MVDR-LBC

Figure 61: A situation where two sources are crossing. One of the sources is more powerful than the others.

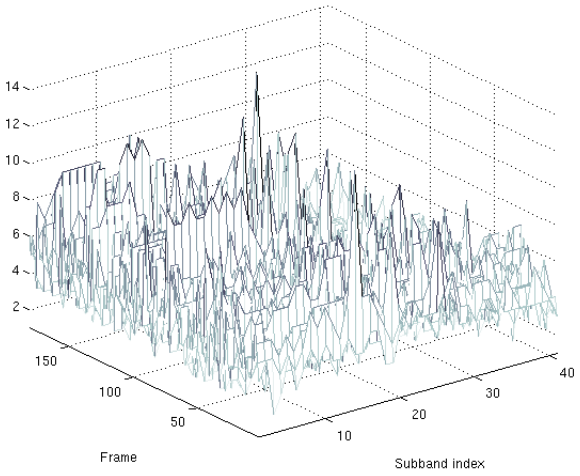


Figure 62: The estimate for the number of sources in MUSIC-S.

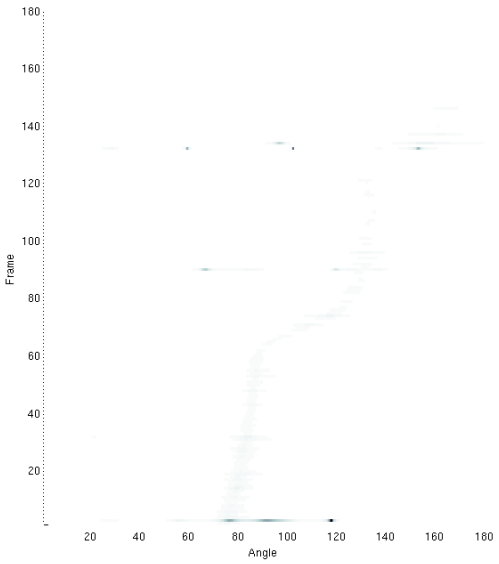


Figure 63: The spectrogram of MUSIC-S without the restriction.

8.2.1 Sparse modeling for the real data

Some real data was compressed using the GRLS algorithm and Golomb-Rice entropy coding described in Section 6. The results are compared to using the exact same algorithms but without the sparsity generating step i.e. OOMP which is a part of GRLS.

For each file 30 seconds i.e. $30 \cdot 8000$ samples were processed from the first channel. The maximum order was 60 and the total length of the sparse predictor was 500.

The compression ratio is defined as

$$\text{Compression ratio} = \frac{\text{Original file size}}{\text{Encoded file size}}$$

and since each sample takes two bytes the Original file size is $30 \cdot 8000 \cdot 2 = 480\,\text{kB}$. The resulting compression ratios are given in Table 1.

file index	**sparse**	**non-sparse**
1	2.096	2.087
2	2.102	2.091
3	2.157	2.131
4	2.096	2.088
5	2.172	2.148
6	2.160	2.134

Table 1: Compression ratios.

file index	**sparse**	**non-sparse**
1	29	53
2	27	44
3	30	46
4	28	46
5	30	43
6	31	42

Table 2: Average orders.

9 Conclusions

This thesis has covered underwater surveillance signal processing in two aspects: by tracking sources using DOA estimation methods and by sparse modeling the underwater signals.

In Section 7 it was revealed that our choice for the type of sensor array is quite different to what has been seen in the literature. The frequency range for its operation is very wide. Considering the type of source signals encountered in underwater surveillance we can conclude that we should be able to utilize the low frequency information which is the least optimal frequency range for the chosen sensor array.

When it comes to the DOA estimation methods, we notice that MVDR is a simple method to implement and provides robust results: Although the sources might not be distiguishable any more as peaks, the presence of them can still be sensed by the wideness of the lobe, for example. The narrowband spectra provided by MVDR can, however, sometimes be contradicting from one frequency band to another and they suffer from bias. These downsides can be alleviated with the original idea of combining the subband estimates using a likelihood function. We built a method called MVDR-LBC to achieve better performance. Experimental Section 8 shows that MVDR-LBC improves the MVDR-S wideband spectra by enabling the sources to still be distiguishable when the subband decisions are contradicting. It can not compete with MUSIC or the coherent subspace methods simply because the resolution of MVDR is not sufficient.

MVDR-LBC could be improved by developing a scheme for detecting the situation of contradicting subband information. The selection of significant subbands would be a place for development also for other methods. In principle, a similar likelihood strategy to MVDR-LBC could also be applied to other narrowband methods if their statistical properties are known. For example l1-SVD could become usable for the wideband situation this way or MUSIC could resolve sources even better.

Compared to MVDR, MUSIC is slightly more difficult to implement because it requires to computation of the eigenvalue decomposition. It also requires a decision for the number of sources and by choosing it badly the performance of MUSIC degrades. This was discovered to be especially true when processing the real data. It was necessary to set a limit for the decision for the number of sources to have a useful result. At its best, MUSIC has better resolution and accuracy than MVDR.

When it comes to the most novel approach to DOA estimation, the sparse methods, the situation is two-folded: It has been shown in the literature and also in the thesis that there are cases where ℓ_1-SVD can compete or even outperform MUSIC as a narrowband estimation method. Its performance is, however, dependent on the

choice of the regularization parameter which proves to be complicated and unpredictable. In addition, the size of the problem grows very large if l1-SVD is used for the wideband problem directly. Similar tractability issues are apparent also in other applications of the sparse methods and they are typically avoided by using a greedy algorithm instead. We have seen that the greedy algorithms, OMP and OOMP, basically use the most basic DOA estimation method – the Beamformer – as the criterion for selecting a source angle. The greedy iterative nature can not recover from a bad decision, which makes the algorithms somewhat inappropriate for the DOA estimation problem at least if the goal is to resolve close sources.

An application where the greedy sparse algorithm proves very useful is the sparse modeling of the underwater signals which can be used to compress signals. We saw that taking the sparsity into account improves the prediction result for real underwater signals. When there is a need to analyze the signal recorded in the sea later on, the lossless compression can turn out useful indeed. Of course, if a compression method like this would be taken into use, it would assumably be very beneficial to utilize the information from other channels (i.e. sensors). This is even related to DOA estimation in the sense that the delays that correspond to the locations of an intra channel predictor would be related to the estimated DOAs.

The family of coherent wideband DOA estimation methods – especially CSSM and WAVES – proved out to be best when it comes to resolution ability. TOPS does not provide similar performance which is expected since it is originally designed to work in high SNR environment. CSSM and WAVES need initial DOA estimates and their performance is quite sensitive to it. The papers describing the coherent methods imply that a low resolution method is sufficient to give this initialization which is why we used MVDR-S as a simple and robust method for this purpose. Along the way, in the results, we tried MUSIC-S and saw that it improves the result. Possibly MVDR-LBC or some other method would work as well. By simulations and experimenting with the real data it is possible to conclude that the spectra of the coherent methods can sometimes distort the magnitudes which means that possibly a strong source makes others invisible in the spectrum which can be dangerous.

It seems that the methods which were discovered as being the best in the thesis have been invented decades ago. All in all, though, only a little of the field of wideband DOA estimation could be covered in this thesis. Numerous other articles have been published and there is active work in the field. For example, the so called maximum likelihood methods were not covered in this thesis although even some research work was done around them. Many of the publications dealing with other applications assume further models of the source signals but such restrictions hardly apply to signals encountered in underwater surveillance. Yet, having more information about the physical measurement system could lead to a model for the noise sources, for

example, which could be very useful.

References

[1] J. Capon, "High-Resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[2] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop, Griffiths, New York*, pp. 243—-258, 1979. Reprinted in IEEE Transactions on Antennas and Propagation, vol. 34, no. 3, pp. 276–380, March 1986.

[3] P. Helin, B. Dumitrescu, J. Astola, and I. Tabus, "Likelihood based combining of subband estimates for wideband DOA," in *Proc. 8th Intl Symposium on Image and Signal Processing and Analysis, Trieste, Italy.*, September, 2013.

[4] H. Hung and M. Kaveh, "Focussing matrices for coherent signal-subspace processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 8, pp. 1272–1281, 1988.

[5] E. DiClaudio and R. Parisi, "WAVES: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2179–2191, Oct. 2001.

[6] Y. Yoon, L. Kaplan, and J. McClellan, "TOPS: New DOA estimator for wideband signals," *IEEE Transactions on Signal Processing*, vol. 54, pp. 1977–1989, June 2006.

[7] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3010–3022, Aug. 2005.

[8] Z. Tang, G. Blacquiere, and G. Leus, "Aliasing-Free wideband beamforming using sparse signal representation," *IEEE Transactions on Signal Processing*, vol. 59, pp. 3464–3469, July 2011.

[9] A. Gretsistas and M. Plumbley, "An alternating descent algorithm for the off-grid doa estimation problem with sparsity constraints," in *Proc. 20th European Signal Processing Conference*, pp. 874–878, 2012.

[10] J. Yin and T. Chen, "Direction-of-Arrival estimation using a sparse representation of array covariance vectors," *IEEE Transactions on Signal Processing*, vol. 59, pp. 4489–4493, Sept. 2011.

[11] B. Dumitrescu, A. Onose, P. Helin, and I. Tabus, "Greedy Sparse RLS," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2194–2207, 2012.

[12] P. Stoica and R. Moses, *Spectral analysis of signals*. Prentice Hall, 2005.

[13] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 387 – 392, Apr. 1985.

[14] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[15] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Processing Letters*, vol. 9, pp. 137–140, 2002.

[16] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[17] J. A. Tropp, Anna, and C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Inform. Theory*, vol. 53, pp. 4655–4666, 2007.

[18] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Transactions on Signal Processing*, vol. 59, pp. 4489 –4493, Sept. 2011.

[19] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation," *IEEE Transactions on Signal Processing*, vol. 59, pp. 4256 –4270, Sept. 2011.

[20] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Transactions on Acoustics Speech Signal Processing*, vol. ASSP-33, pp. 823–831, Aug. 1985.

[21] J. Krolik, "Focused wideband array processing for spatial spectral estimation," in *Advances in Spectrum Analysis and Array Processing* (S. Haykin, ed.), Englewood Cliffs, NJ: Prentice-Hall, 1991.

[22] D. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," Master's thesis, MIT, July 2003.

[23] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," tech. rep., Department of Electrical and Computer Engineering, Rice University, 2005.

[24] K. Mahata, "A Subspace Algorithm for WideBand Source Localization Without NarrowBand Filtering," *IEEE Transactions on Signal Processing*, vol. 59, pp. 3470–3475, July 2011.

[25] C. Vaidyanathan and K. Buckley, "Performance analysis of the MVDR spatial spectrum estimator," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1427–1437, Jun. 1995.

[26] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 4. ed., 2001.

[27] J. Coalson, "FLAC - Free lossless audio codec." `http://flac.sourceforge.net`.

[28] D. Giacobello, T. van Waterschoot, M. G. Christensen, S. H. Jensen, and M. Moonen, "High-order sparse linear predictors for audio processing," in *Proc. 18th European Signal Processing Conference Aalborg, Denmark*, Aug., 2010.

[29] F. Ghido and I. Tabus, "Sparse modeling for lossless audio compression," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 1, pp. 14–28, 2013.

[30] H. Huang, S. Rahardja, X. Lin, R. Yu, and P. Franti, "Cascaded RLS-LMS prediction in MPEG-4 lossless audio coding," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. V–V, 2006.

[31] S. Golomb, "Run-length encodings," *IEEE Transactions on Information Theory*, vol. 12, no. 3, pp. 399–401, 1966.

[32] R. F. Rice and R. Plaunt, "Adaptive variable-length coding for efficient compression of spacecraft television data," *IEEE Transactions on Communications*, vol. 16, no. 9, pp. 889—-897, 1971.

[33] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. V – 337–40 vol.5, april 2003.

[34] W. S. Burdic, *Underwater acoustic system analysis*. Prentice Hall, 1991.

[35] M. F. McKenna, D. Ross, S. M. Wiggins, and J. A. Hildebrand, "Underwater radiated noise from modern commercial ships," *J. Acoust. Soc. Am.*, vol. 131, pp. 92–103, 2012.